Contents

AIM-INSTALLATION OF HADOOP

STEPS-

1. Prerequisites

First, we need to make sure that the following prerequisites are installed:

1. Java 8 runtime environment (JRE): Hadoop 3 requires a Java 8 installation. I prefer using the offline installer.

2. Java 8 development Kit (JDK)

3. To unzip downloaded Hadoop binaries, we should install 7zip.

4. I will create a folder “E:\hadoop-env” on my local machine to store downloaded files.

2. Download Hadoop binaries

The first step is to download Hadoop binaries from the official website. The binary package size is about 342 MB.

Image for post

Figure 1 — Hadoop binaries download link

After finishing the file download, we should unpack the package using 7zip int two steps. First, we should extract the hadoop-3.2.1.tar.gz library, and then, we should unpack the extracted tar file:

Image for post

Figure 2 — Extracting hadoop-3.2.1.tar.gz package using 7zip

Image for post

Figure 3 — Extracted hadoop-3.2.1.tar file

Image for post

Figure 4 — Extracting the hadoop-3.2.1.tar file

The tar file extraction may take some minutes to finish. In the end, you may see some warnings about symbolic link creation. Just ignore these warnings since they are not related to windows.

Image for post

Figure 5 — Symbolic link warnings

After unpacking the package, we should add the Hadoop native IO libraries, which can be found in the following GitHub repository: https://github.com/cdarlint/winutils.

Since we are installing Hadoop 3.2.1, we should download the files located in https://github.com/cdarlint/winutils/tree/master/hadoop-3.2.1/bin and copy them into the “hadoop-3.2.1\bin” directory.

3. Setting up environment variables

After installing Hadoop and its prerequisites, we should configure the environment variables to define Hadoop and Java default paths.

To edit environment variables, go to Control Panel > System and Security > System (or right-click > properties on My Computer icon) and click on the “Advanced system settings” link.

Image for post

Figure 6 — Opening advanced system settings

When the “Advanced system settings” dialog appears, go to the “Advanced” tab and click on the “Environment variables” button located on the bottom of the dialog.

Image for post

Figure 7 — Advanced system settings dialog

In the “Environment Variables” dialog, press the “New” button to add a new variable.

Note: In this guide, we will add user variables since we are configuring Hadoop for a single user. If you are looking to configure Hadoop for multiple users, you can define System variables instead.

There are two variables to define:

1. JAVA_HOME: JDK installation folder path

2. HADOOP_HOME: Hadoop installation folder path

Image for post

Figure 8 — Adding JAVA_HOME variable

Image for post

Figure 9 — Adding HADOOP_HOME variable

Now, we should edit the PATH variable to add the Java and Hadoop binaries paths as shown in the following screenshots.

Image for post

Figure 10 — Editing the PATH variable

Image for post

Figure 11 — Editing PATH variable

Image for post

Figure 12— Adding new paths to the PATH variable

3.1. JAVA_HOME is incorrectly set error

Now, let’s open PowerShell and try to run the following command:

hadoop -version

In this example, since the JAVA_HOME path contains spaces, I received the following error:

JAVA_HOME is incorrectly set

Image for post

Figure 13 — JAVA_HOME error

To solve this issue, we should use the windows 8.3 path instead. As an example:

Use “Progra~1” instead of “Program Files”

Use “Progra~2” instead of “Program Files(x86)”

After replacing “Program Files” with “Progra~1”, we closed and reopened PowerShell and tried the same command. As shown in the screenshot below, it runs without errors.

Image for post

Figure 14 — hadoop -version command executed successfully

4. Configuring Hadoop cluster

There are four files we should alter to configure Hadoop cluster:

%HADOOP_HOME%\etc\hadoop\hdfs-site.xml

%HADOOP_HOME%\etc\hadoop\core-site.xml

%HADOOP_HOME%\etc\hadoop\mapred-site.xml

%HADOOP_HOME%\etc\hadoop\yarn-site.xml

4.1. HDFS site configuration

As we know, Hadoop is built using a master-slave paradigm. Before altering the HDFS configuration file, we should create a directory to store all master node (name node) data and another one to store data (data node). In this example, we created the following directories:

E:\hadoop-env\hadoop-3.2.1\data\dfs\namenode

E:\hadoop-env\hadoop-3.2.1\data\dfs\datanode

Now, let’s open “hdfs-site.xml” file located in “%HADOOP_HOME%\etc\hadoop” directory, and we should add the following properties within the <configuration></configuration> element:

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<name>dfs.namenode.name.dir</name>

<value>file:///E:/hadoop-env/hadoop-3.2.1/data/dfs/namenode</value>

</property>

<property>

<name>dfs.datanode.data.dir</name>

<value>file:///E:/hadoop-env/hadoop-3.2.1/data/dfs/datanode</value>

</property>

Note that we have set the replication factor to 1 since we are creating a single node cluster.

4.2. Core site configuration

Now, we should configure the name node URL adding the following XML code into the <configuration></configuration> element within “core-site.xml”:

<property>

<name>fs.default.name</name>

<value>hdfs://localhost:9820</value>

</property>

4.3. Map Reduce site configuration

Now, we should add the following XML code into the <configuration></configuration> element within “mapred-site.xml”:

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

<description>MapReduce framework name</description>

</property>

4.4. Yarn site configuration

Now, we should add the following XML code into the <configuration></configuration> element within “yarn-site.xml”:

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce_shuffle</value>

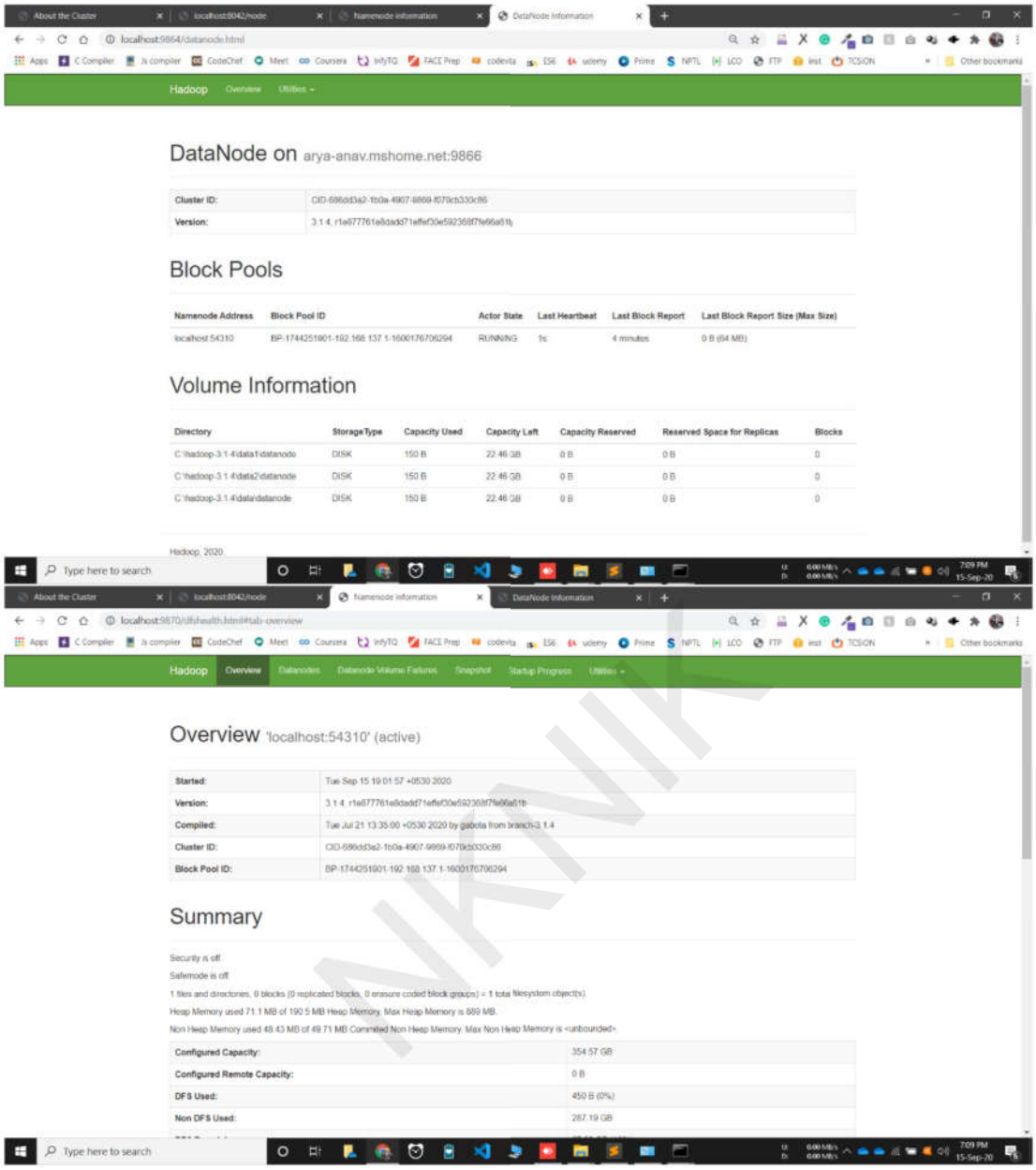
<description>Yarn Node Manager Aux Service</description>

</property>

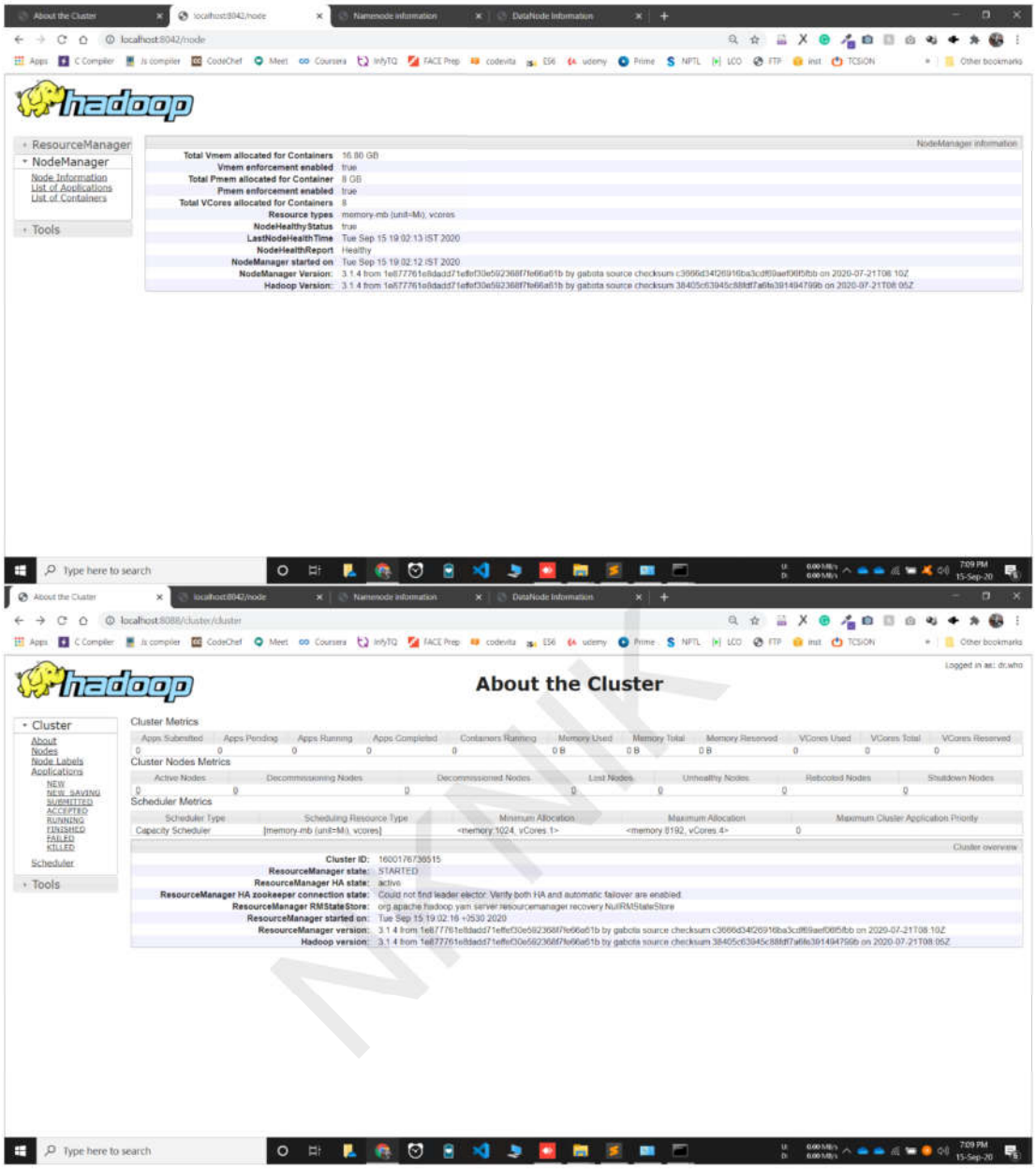
5. Formatting Name node

After finishing the configuration, let’s try to format the name node using the following command:

hdfs namenode -format


About the Cluster
localhost:8042/node
Namenode information
DataNode Information
localhost:9864/datanode.html
Hadoop
Overview
Utilities
DataNode on arya-anav.mshome.net:9866
Cluster ID: CID-686dd3a2-1b0a-4907-9869-f070cb330c86
Version: 3.1.4, r1e877761e8dadd71effef30e592368ff7e66a61b
Block Pools
Namenode Address
Block Pool ID
Actor State
Last Heartbeat
Last Block Report
Last Block Report Size (Max Size)
localhost:54310
BP-1744251901-192.168.137.1-1600176706294
RUNNING
1s
4 minutes
0 B (64 MB)
Volume Information
Directory
StorageType
Capacity Used
Capacity Left
Capacity Reserved
Reserved Space for Replicas
Blocks
C:\hadoop-3.1.4\data1\datanode
C:\hadoop-3.1.4\data2\datanode
C:\hadoop-3.1.4\data\datanode
DISK
150 B
22.46 GB
0 B
Hadoop, 2020.
Type here to search
7:09 PM
15-Sep-20

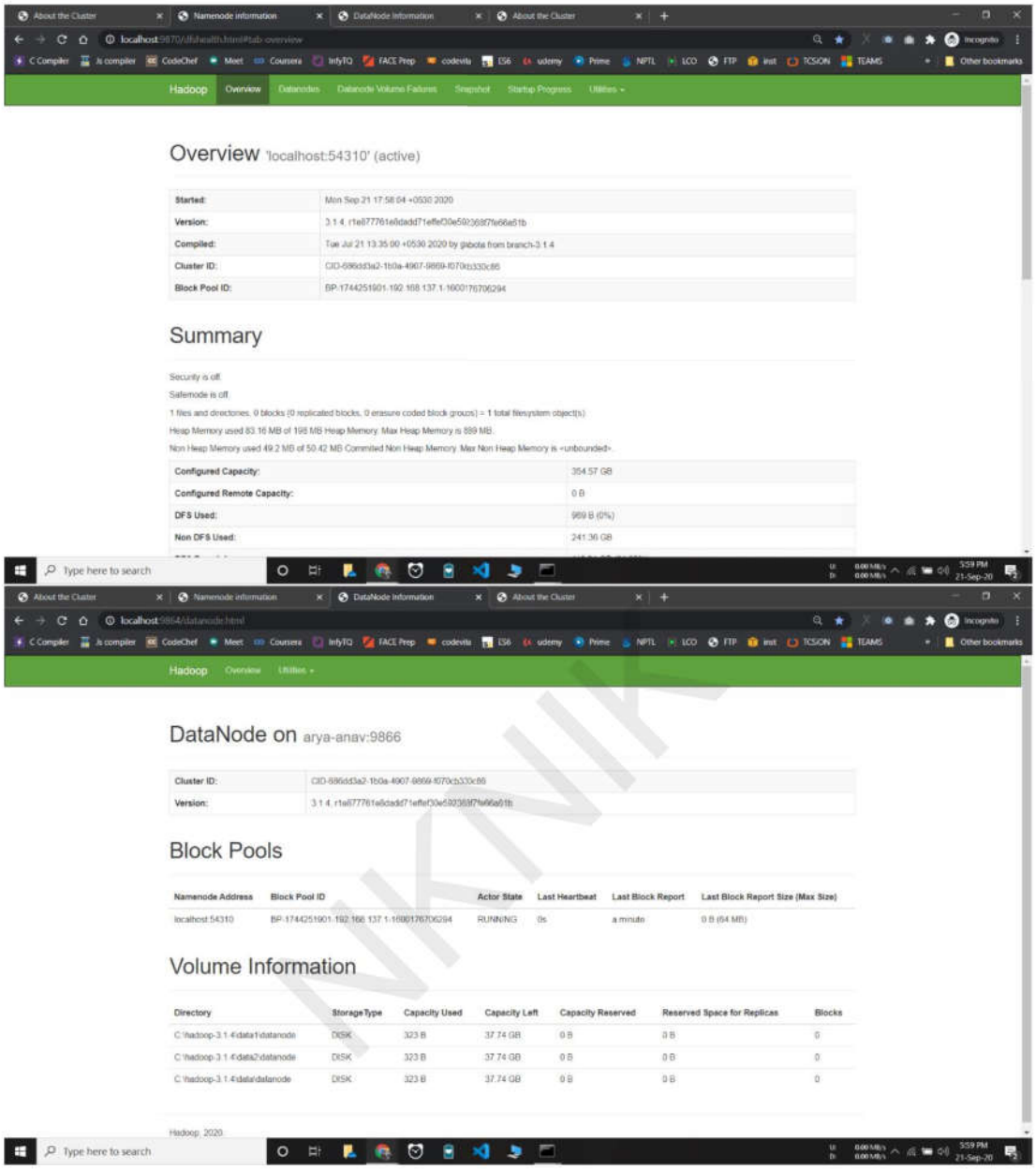
About the Cluster
localhost:8042/node
Namenode information
DataNode Information
localhost:9870/dfshealth.html#tab-overview
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Overview 'localhost:54310' (active)
Started: Tue Sep 15 19:01:57 +0530 2020
Version: 3.1.4, r1e877761e8dadd71effef30e592368ff7e66a61b
Compiled: Tue Jul 21 13:35:00 +0530 2020 by gabota from branch-3.1.4
Cluster ID: CID-686dd3a2-1b0a-4907-9869-f070cb330c86
Block Pool ID: BP-1744251901-192.168.137.1-1600176706294
Summary
Security is off.
Safemode is off.
1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).
Heap Memory used 71.1 MB of 190.5 MB Heap Memory. Max Heap Memory is 889 MB.
Non Heap Memory used 48.43 MB of 49.71 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.
Configured Capacity: 354.57 GB
Configured Remote Capacity: 0 B
DFS Used: 450 B (0%)
Non DFS Used: 287.19 GB
Type here to search
7:09 PM
15-Sep-20


About the Cluster
localhost:8042/node
Namenode information
DataNode Information
localhost:8042/node
Apps
C Compiler
Js compiler
CodeChef
Meet
Coursera
InfyTQ
FACE Prep
codevita
ES6
udemy
Prime
NPTL
LCO
FTP
inst
TCSION
Other bookmarks
hadoop
ResourceManager
NodeManager
Node Information
List of Applications
List of Containers
Tools
NodeManager information
Total Vmem allocated for Containers 16.80 GB
Vmem enforcement enabled true
Total Pmem allocated for Container 8 GB
Pmem enforcement enabled true
Total VCores allocated for Containers 8
Resource types memory-mb (unit=Mi), vcores
NodeHealthy Status true
LastNodeHealthTime Tue Sep 15 19:02:13 IST 2020
NodeHealthReport Healthy
NodeManager started on Tue Sep 15 19:02:12 IST 2020
NodeManager Version: 3.1.4 from 1e877761e8dadd71effef30e592368f7fe66a61b by gabota source checksum c3666d3426916ba3cdf89aef06f5fbb on 2020-07-21T08:10Z
Hadoop Version: 3.1.4 from 1e877761e8dadd71effef30e592368f7fe66a61b by gabota source checksum 38405c63945c88fdf7a9fe391494799b on 2020-07-21T08:05Z
Type here to search
7:09 PM
15-Sep-20

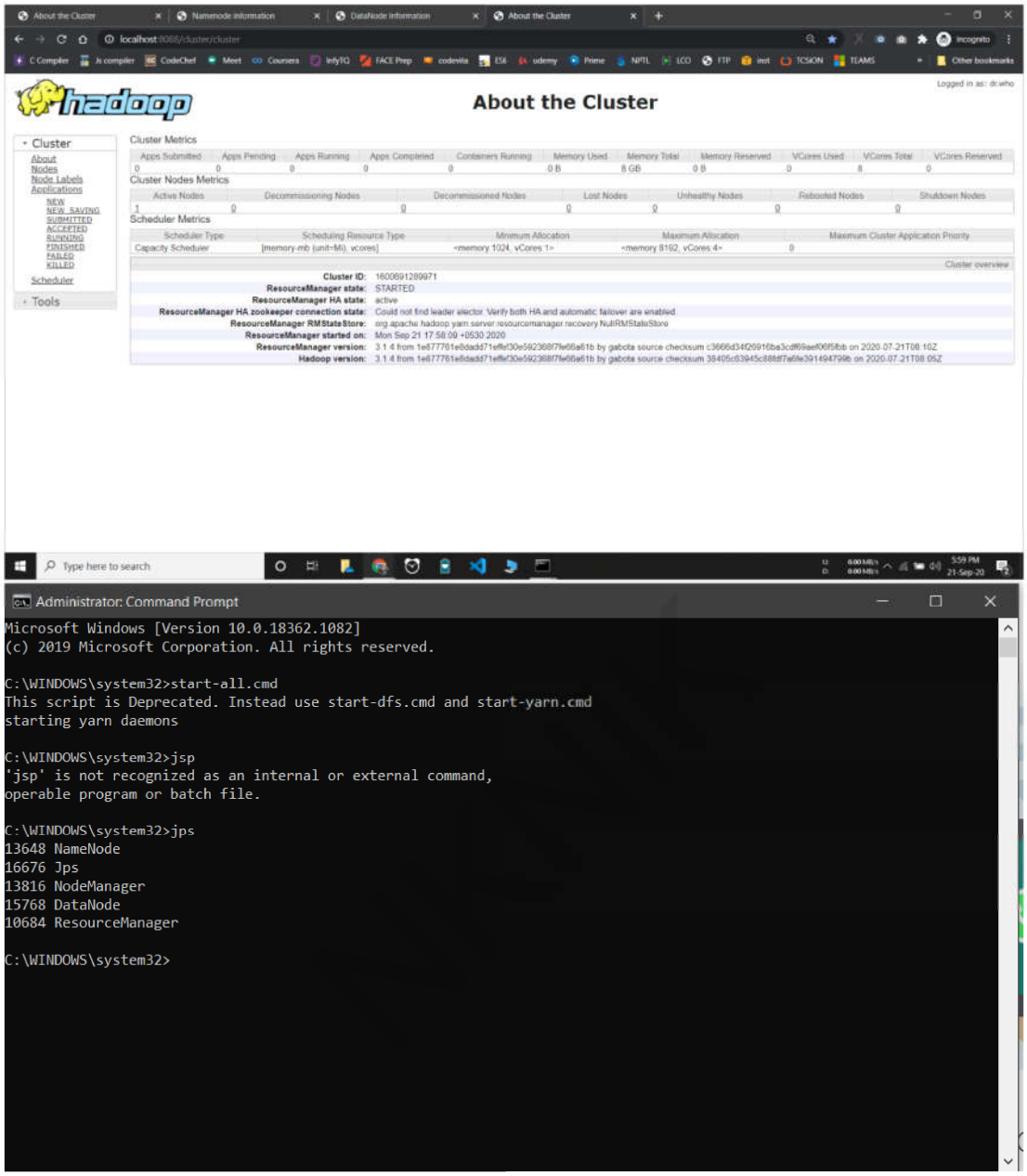
About the Cluster
localhost:8088/cluster/cluster
Apps
C Compiler
Js compiler
CodeChef
Meet
Coursera
InfyTQ
FACE Prep
codevita
ES6
udemy
Prime
NPTL
LCO
FTP
inst
TCSION
Other bookmarks
hadoop
Logged in as: dr.who
About the Cluster
Cluster
About
Nodes
Node Labels
Applications
NEW
NEW_SAVING
SUBMITTED
ACCEPTED
RUNNING
FINISHED
FAILED
KILLED
Scheduler
Tools
Cluster Metrics
Apps Submitted
Apps Pending
Apps Running
Apps Completed
Containers Running
Memory Used
Memory Total
Memory Reserved
VCores Used
VCores Total
VCores Reserved
0 0 0 0 0 0 B 0 B 0 B 0 0 0
Cluster Nodes Metrics
Active Nodes
Decommissioning Nodes
Decommissioned Nodes
Lost Nodes
Unhealthy Nodes
Rebooted Nodes
Shutdown Nodes
0 0 0 0 0 0 0
Scheduler Metrics
Scheduler Type
Scheduling Resource Type
Minimum Allocation
Maximum Allocation
Maximum Cluster Application Priority
Capacity Scheduler [memory-mb (unit=Mi), vcores] <memory:1024, vCores:1> <memory:8192, vCores:4> 0
Cluster overview
Cluster ID: 1600176736515
ResourceManager state: STARTED
ResourceManager HA state: active
ResourceManager HA zookeeper connection state: Could not find leader elector. Verify both HA and automatic failover are enabled.
ResourceManager RMStateStore: org.apache.hadoop.yarn.server.resourcemanager.recovery.NullRMStateStore
ResourceManager started on: Tue Sep 15 19:02:16 +0530 2020
ResourceManager version: 3.1.4 from 1e877761e8dadd71effef30e592368f7fe66a61b by gabota source checksum c3666d3426916ba3cdf89aef06f5fbb on 2020-07-21T08:10Z
Hadoop version: 3.1.4 from 1e877761e8dadd71effef30e592368f7fe66a61b by gabota source checksum 38405c63945c88fdf7a9fe391494799b on 2020-07-21T08:05Z
Type here to search
7:09 PM
15-Sep-20


Apache Hadoop Distribution - yarn nodemanager
Apache Hadoop Distribution - yarn resourcemanager
Apache Hadoop Distribution - hadoop datanode
Apache Hadoop Distribution - hadoop namenode
Administrator: Command Prompt
Turned in
pract1.pdf
Unsubmit
comments

About the Cluster
Logged in as: dr.who
Cluster
About
Nodes
Node Labels
Applications
NEW
NEW_SAVING
SUBMITTED
ACCEPTED
RUNNING
FINISHED
FAILED
KILLED
Scheduler
Tools
Cluster Metrics
Apps Submitted
Apps Pending
Apps Running
Apps Completed
Containers Running
Memory Used
Memory Total
Memory Reserved
VCores Used
VCores Total
VCores Reserved
Cluster Nodes Metrics
Active Nodes
Decommissioning Nodes
Decommissioned Nodes
Lost Nodes
Unhealthy Nodes
Rebooted Nodes
Shutdown Nodes
Scheduler Metrics
Scheduler Type
Scheduling Resource Type
Minimum Allocation
Maximum Allocation
Maximum Cluster Application Priority
Capacity Scheduler
[memory-mb (unit=Mi), vcores]
<memory:1024, vCores:1>
<memory:8192, vCores:4>
Cluster overview
Cluster ID: 1600891289971
ResourceManager state: STARTED
ResourceManager HA state: active
ResourceManager HA zookeeper connection state: Could not find leader elector. Verify both HA and automatic failover are enabled.
ResourceManager RMStateStore: org.apache.hadoop.yarn.server.resourcemanager.recovery.NullRMStateStore
ResourceManager started on: Mon Sep 21 17:58:09 +0530 2020
ResourceManager version: 3.1.4 from 1e877761e8dadd71ef8e30e592368f71e6fa61fb by gabota source checksum c3666d34f25916ba3cdf69aef0665fbb on 2020-07-21T08:10Z
Hadoop version: 3.1.4 from 1e877761e8dadd71ef8e30e592368f71e6fa61fb by gabota source checksum 38405c53945c88fdf7a66e3914947996 on 2020-07-21T08:05Z


About the Cluster
Namenode information
DataNode Information
About the Cluster
localhost:9870/dfshealth.html#tab-overview
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Overview 'localhost:54310' (active)
Started: Mon Sep 21 17:58:04 +0530 2020
Version: 3.1.4, r1e877761e6dadd71effef30e5923681f7e66e51b
Compiled: Tue Jul 21 13:35:00 +0530 2020 by gabota from branch-3.1.4
Cluster ID: CID-686dd3a2-1b0a-4907-9869-f070cb330c86
Block Pool ID: BP-1744251901-192.168.137.1-1600176706294
Summary
Security is off.
Safemode is off.
1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).
Heap Memory used 83.16 MB of 198 MB Heap Memory. Max Heap Memory is 889 MB.
Non Heap Memory used 49.2 MB of 50.42 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.
Configured Capacity: 354.57 GB
Configured Remote Capacity: 0 B
DFS Used: 969 B (0%)
Non DFS Used: 241.36 GB
Type here to search
5:59 PM
21-Sep-20

About the Cluster
Namenode information
DataNode Information
About the Cluster
localhost:9864/datanode.html
Hadoop
Overview
Utilities
DataNode on arya-anav:9866
Cluster ID: CID-686dd3a2-1b0a-4907-9869-f070cb330c86
Version: 3.1.4, r1e877761e6dadd71effef30e5923681f7e66e51b
Block Pools
Namenode Address
Block Pool ID
Actor State
Last Heartbeat
Last Block Report
Last Block Report Size (Max Size)
localhost:54310
BP-1744251901-192.168.137.1-1600176706294
RUNNING
0s
a minute
0 B (64 MB)
Volume Information
Directory
StorageType
Capacity Used
Capacity Left
Capacity Reserved
Reserved Space for Replicas
Blocks
C:\hadoop-3.1.4\data1\datanode
DISK
323 B
37.74 GB
0 B
0 B
0
C:\hadoop-3.1.4\data2\datanode
DISK
323 B
37.74 GB
0 B
0 B
0
C:\hadoop-3.1.4\data\datanode
DISK
323 B
37.74 GB
0 B
0 B
0
Hadoop, 2020.
Type here to search
5:59 PM
21-Sep-20


About the Cluster
Namenode information
DataNode information
About the Cluster
localhost:8088/cluster/cluster
Incognito
Other bookmarks
Logged in as: dr.who
About the Cluster
Cluster
About
Nodes
Node Labels
Applications
NEW
NEW_SAVING
SUBMITTED
ACCEPTED
RUNNING
FINISHED
FAILED
KILLED
Scheduler
Tools
Cluster Metrics
Apps Submitted
Apps Pending
Apps Running
Apps Completed
Containers Running
Memory Used
Memory Total
Memory Reserved
VCores Used
VCores Total
VCores Reserved
0 B
8 GB
0 B
8
Cluster Nodes Metrics
Active Nodes
Decommissioning Nodes
Decommissioned Nodes
Lost Nodes
Unhealthy Nodes
Rebooted Nodes
Shutdown Nodes
Scheduler Metrics
Scheduler Type
Scheduling Resource Type
Minimum Allocation
Maximum Allocation
Maximum Cluster Application Priority
Capacity Scheduler
[memory-mb (unit=Mi), vcores]
<memory:1024, vCores:1>
<memory:8192, vCores:4>
Cluster overview
Cluster ID: 1600691289971
ResourceManager state: STARTED
ResourceManager HA state: active
ResourceManager HA zookeeper connection state: Could not find leader elector. Verify both HA and automatic failover are enabled
ResourceManager RMStateStore: org.apache.hadoop.yarn.server.resourcemanager.recovery.NullRMStateStore
ResourceManager started on: Mon Sep 21 17:58:09 +0530 2020
ResourceManager version: 3.1.4 from 1e877761e8dadd71effef30e592368f7fe66a61b by gabota source checksum c3666d34f20916ba3cdf69aef06f5fdb on 2020-07-21T08:10Z
Hadoop version: 3.1.4 from 1e877761e8dadd71effef30e592368f7fe66a61b by gabota source checksum 38405c63945c88fdf7af6fe391494799b on 2020-07-21T08:05Z
Type here to search
5:59 PM
21-Sep-20

Administrator: Command Prompt
Microsoft Windows [Version 10.0.18362.1082]
(c) 2019 Microsoft Corporation. All rights reserved.
C:\WINDOWS\system32>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons
C:\WINDOWS\system32>jsp
'jsp' is not recognized as an internal or external command,
operable program or batch file.
C:\WINDOWS\system32>jps
13648 NameNode
16676 Jps
13816 NodeManager
15768 DataNode
10684 ResourceManager
C:\WINDOWS\system32>

AIM-HDFS commands

STEPS-

ls: This command is used to list all the files. Use lsr for recursive approach. It is useful when we want a hierarchy of a folder.

Syntax:

ADVERTISING

bin/hdfs dfs -ls <path>

Example:

bin/hdfs dfs -ls /

It will print all the directories present in HDFS. bin directory contains executables so, bin/hdfs means we want the executables of hdfs particularly dfs(Distributed File System) commands.

mkdir: To create a directory. In Hadoop dfs there is no home directory by default. So let's first create it.

Syntax:

bin/hdfs dfs -mkdir <folder name>

creating home directory:

hdfs/bin -mkdir /user

hdfs/bin -mkdir /user/username -> write the username of your computer

Example:

bin/hdfs dfs -mkdir /NK => '/' means absolute path

bin/hdfs dfs -mkdir NK2 => Relative path -> the folder will be

created relative to the home directory.

touchz: It creates an empty file.

Syntax:

bin/hdfs dfs -touchz <file_path>

Example:

bin/hdfs dfs -touchz /NK/myfile.txt

copyFromLocal (or) put: To copy files/folders from local file system to hdfs store. This is the most important command. Local filesystem means the files present on the OS.

Syntax:

bin/hdfs dfs -copyFromLocal <local file path> <dest(present on hdfs)>

Example: Let's suppose we have a file AI.txt on Desktop which we want to copy to folder NK present on hdfs.

bin/hdfs dfs -copyFromLocal ../Desktop/AI.txt /NK

(OR)

bin/hdfs dfs -put ../Desktop/AI.txt /NK

cat: To print file contents.

Syntax:

bin/hdfs dfs -cat <path>

Example:

// print the content of AI.txt present

// inside NK folder.

bin/hdfs dfs -cat /NK/AI.txt ->

copyToLocal (or) get: To copy files/folders from hdfs store to local file system.

Syntax:

bin/hdfs dfs -copyToLocal <<srcfile(on hdfs)> <local file dest>

Example:

bin/hdfs dfs -copyToLocal /NK ../Desktop/hero

(OR)

bin/hdfs dfs -get /NK/myfile.txt ../Desktop/hero

myfile.txt from NK folder will be copied to folder hero present on Desktop.

Note: Observe that we don't write bin/hdfs while checking the things present on local filesystem.

moveFromLocal: This command will move file from local to hdfs.

Syntax:

bin/hdfs dfs -moveFromLocal <local src> <dest(on hdfs)>

Example:

bin/hdfs dfs -moveFromLocal ../Desktop/cutAndPaste.txt /NK

cp: This command is used to copy files within hdfs. Lets copy folder NK to NK_copied.

Syntax:

bin/hdfs dfs -cp <src(on hdfs)> <dest(on hdfs)>

Example:

bin/hdfs -cp /NK /NK_copied

mv: This command is used to move files within hdfs. Lets cut-paste a file myfile.txt from NK folder to NK_copied.

Syntax:

bin/hdfs dfs -mv <src(on hdfs)> <src(on hdfs)>

Example:

bin/hdfs -mv /NK/myfile.txt /NK_copied

rmr: This command deletes a file from HDFS recursively. It is very useful command when you want to delete a non-empty directory.

Syntax:

bin/hdfs dfs -rmr <filename/directoryName>

Example:

bin/hdfs dfs -rmr /NK_copied -> It will delete all the content inside the

directory then the directory itself.

du: It will give the size of each file in directory.

Syntax:

bin/hdfs dfs -du <dirName>

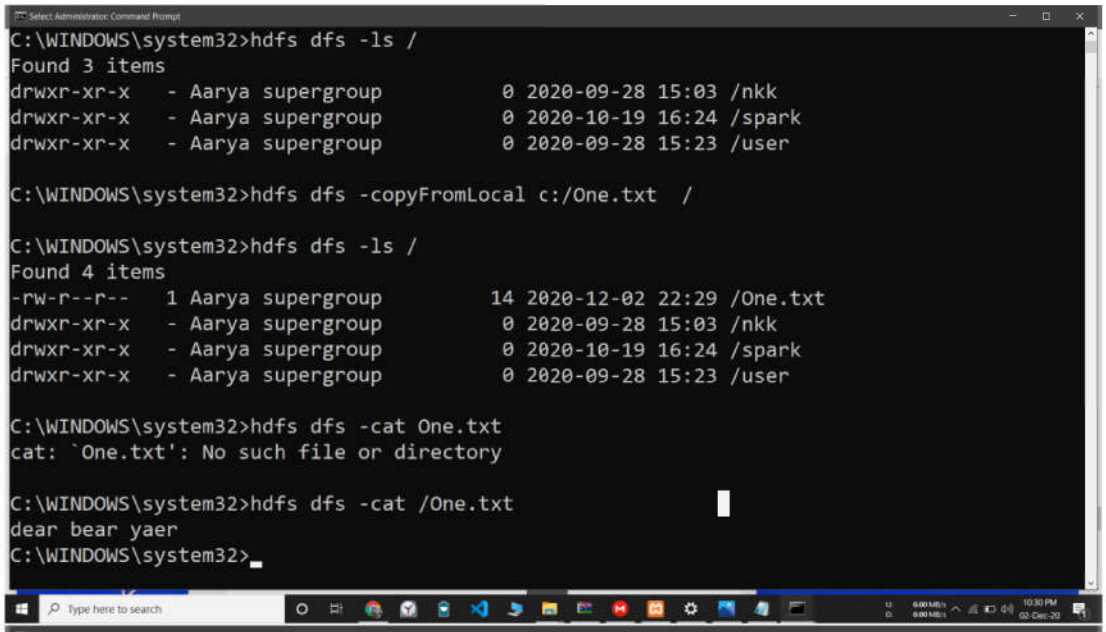
Example:

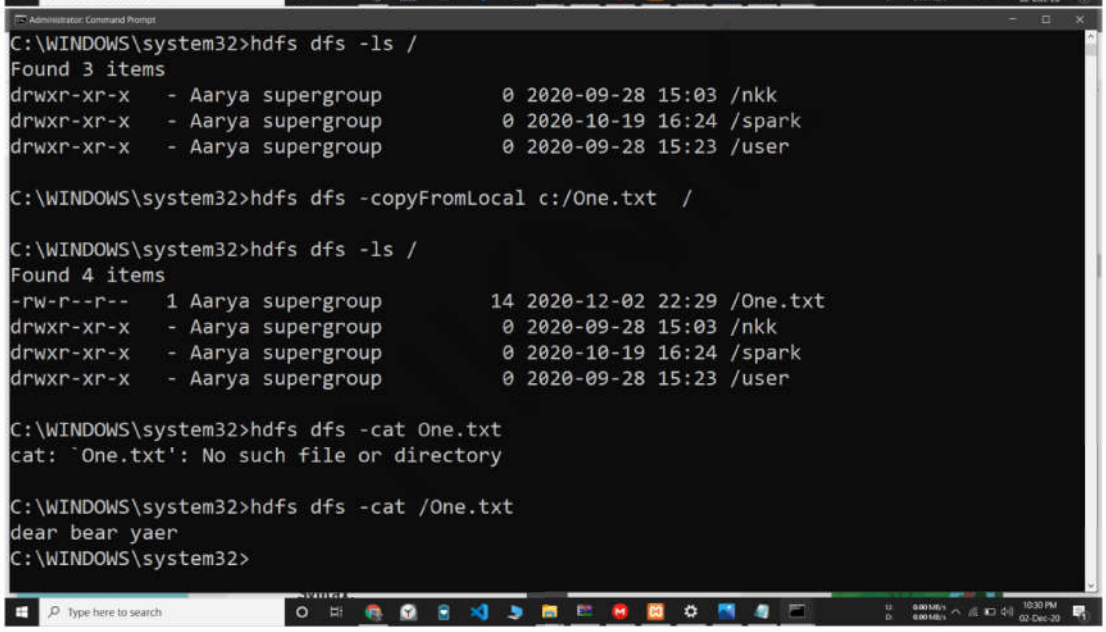
bin/hdfs dfs -du /NK

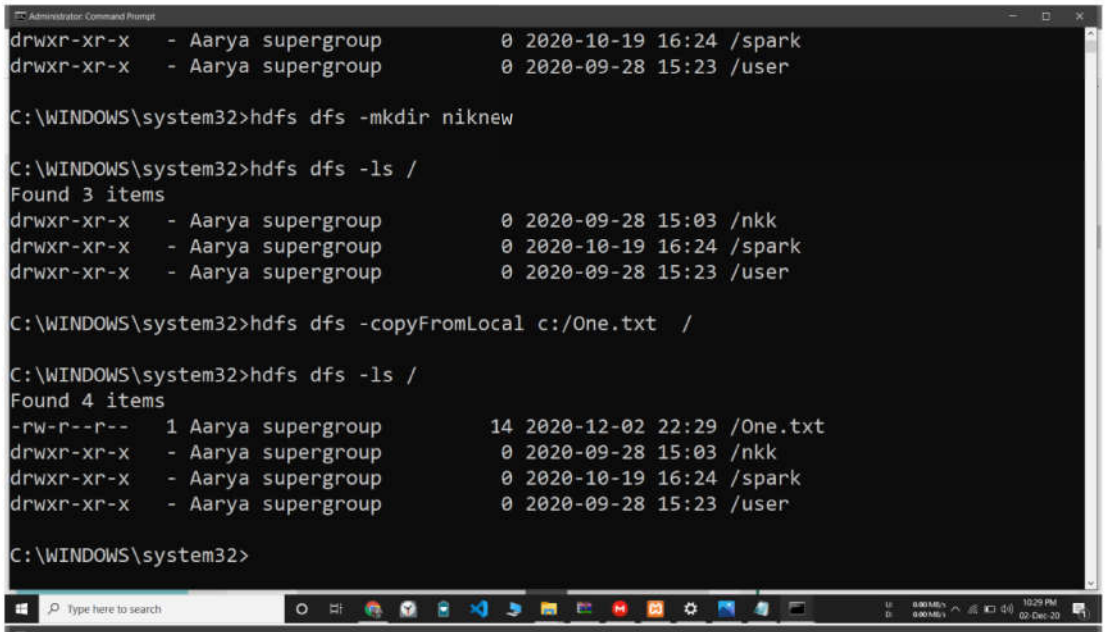
dus:: This command will give the total size of directory/file.

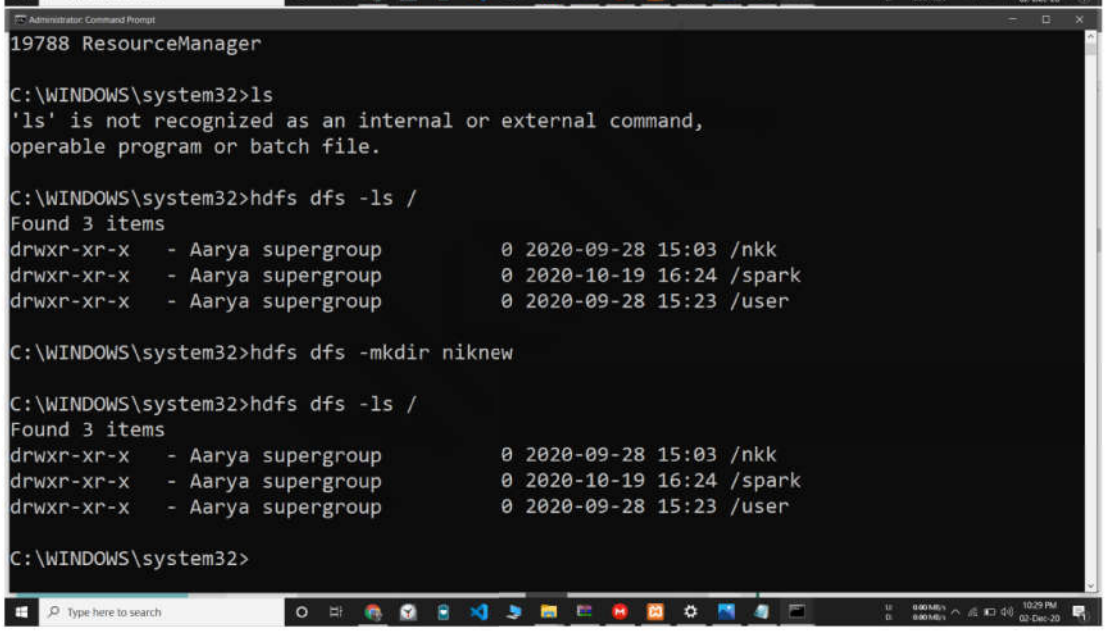
Syntax:

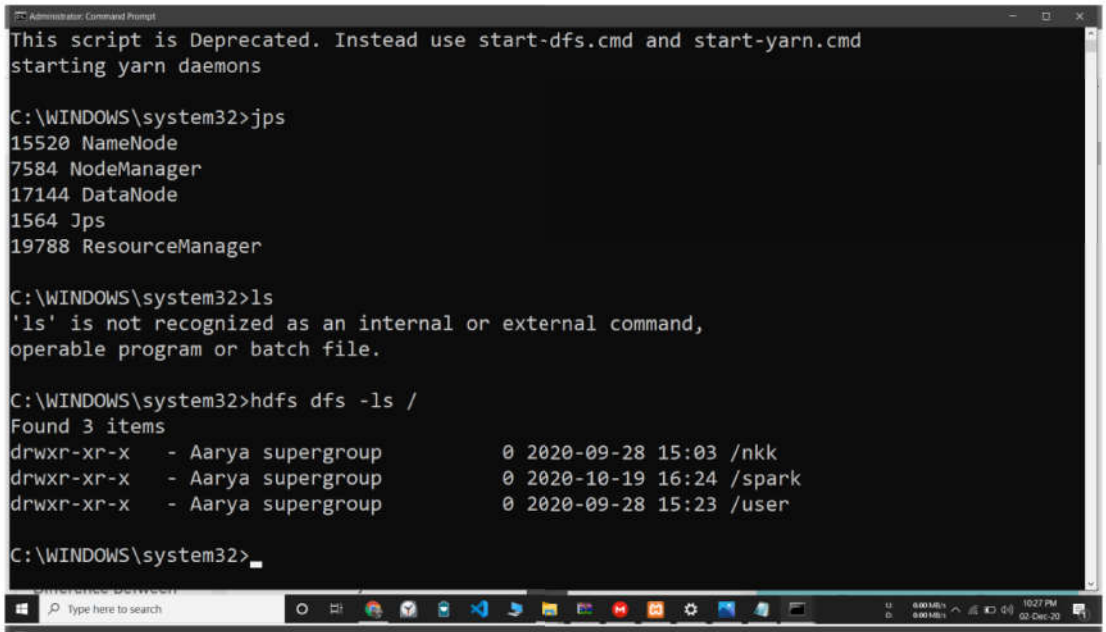
bin/hdfs dfs -dus <dirName>

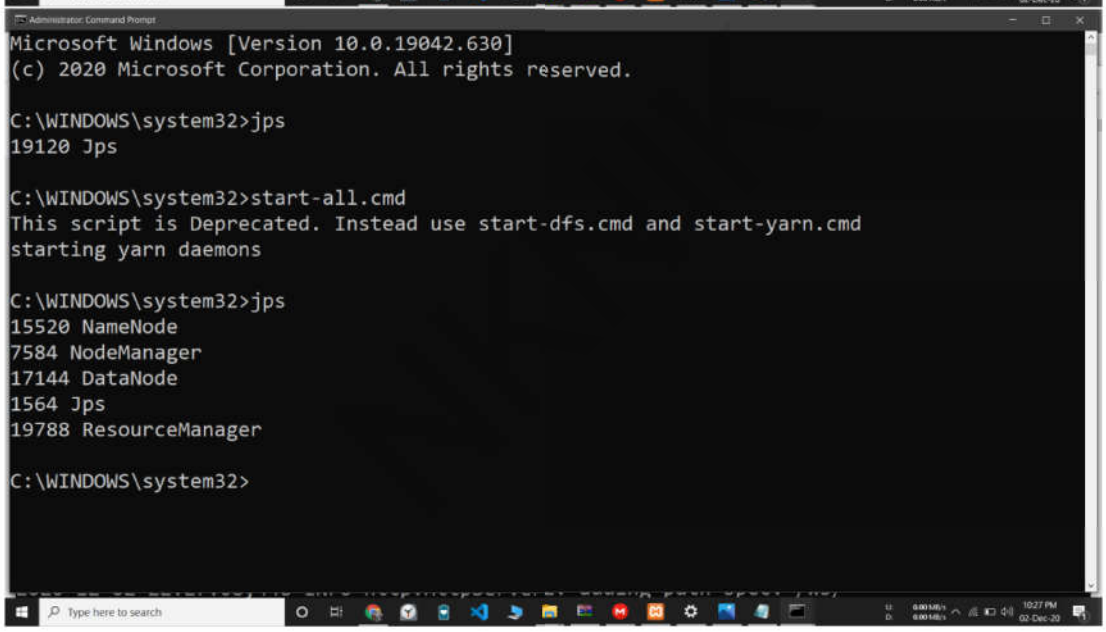

Select Administrator: Command Prompt
C:\WINDOWS\system32>hdfs dfs -ls /
Found 3 items
drwxr-xr-x - Aarya supergroup 0 2020-09-28 15:03 /nkk
drwxr-xr-x - Aarya supergroup 0 2020-10-19 16:24 /spark
drwxr-xr-x - Aarya supergroup 0 2020-09-28 15:23 /user
C:\WINDOWS\system32>hdfs dfs -copyFromLocal c:/One.txt /
C:\WINDOWS\system32>hdfs dfs -ls /
Found 4 items
-rw-r--r-- 1 Aarya supergroup 14 2020-12-02 22:29 /One.txt
drwxr-xr-x - Aarya supergroup 0 2020-09-28 15:03 /nkk
drwxr-xr-x - Aarya supergroup 0 2020-10-19 16:24 /spark
drwxr-xr-x - Aarya supergroup 0 2020-09-28 15:23 /user
C:\WINDOWS\system32>hdfs dfs -cat One.txt
cat: `One.txt': No such file or directory
C:\WINDOWS\system32>hdfs dfs -cat /One.txt
dear bear yaer
C:\WINDOWS\system32>
Type here to search


Administrator: Command Prompt
C:\WINDOWS\system32>hdfs dfs -ls /
Found 3 items
drwxr-xr-x - Aarya supergroup 0 2020-09-28 15:03 /nkk
drwxr-xr-x - Aarya supergroup 0 2020-10-19 16:24 /spark
drwxr-xr-x - Aarya supergroup 0 2020-09-28 15:23 /user
C:\WINDOWS\system32>hdfs dfs -copyFromLocal c:/One.txt /
C:\WINDOWS\system32>hdfs dfs -ls /
Found 4 items
-rw-r--r-- 1 Aarya supergroup 14 2020-12-02 22:29 /One.txt
drwxr-xr-x - Aarya supergroup 0 2020-09-28 15:03 /nkk
drwxr-xr-x - Aarya supergroup 0 2020-10-19 16:24 /spark
drwxr-xr-x - Aarya supergroup 0 2020-09-28 15:23 /user
C:\WINDOWS\system32>hdfs dfs -cat One.txt
cat: `One.txt': No such file or directory
C:\WINDOWS\system32>hdfs dfs -cat /One.txt
dear bear yaer
C:\WINDOWS\system32>
Type here to search


Administrator: Command Prompt
drwxr-xr-x - Aarya supergroup 0 2020-10-19 16:24 /spark
drwxr-xr-x - Aarya supergroup 0 2020-09-28 15:23 /user
C:\WINDOWS\system32>hdfs dfs -mkdir niknew
C:\WINDOWS\system32>hdfs dfs -ls /
Found 3 items
drwxr-xr-x - Aarya supergroup 0 2020-09-28 15:03 /nkk
drwxr-xr-x - Aarya supergroup 0 2020-10-19 16:24 /spark
drwxr-xr-x - Aarya supergroup 0 2020-09-28 15:23 /user
C:\WINDOWS\system32>hdfs dfs -copyFromLocal c:/One.txt /
C:\WINDOWS\system32>hdfs dfs -ls /
Found 4 items
-rw-r--r-- 1 Aarya supergroup 14 2020-12-02 22:29 /One.txt
drwxr-xr-x - Aarya supergroup 0 2020-09-28 15:03 /nkk
drwxr-xr-x - Aarya supergroup 0 2020-10-19 16:24 /spark
drwxr-xr-x - Aarya supergroup 0 2020-09-28 15:23 /user
C:\WINDOWS\system32>
Type here to search


Administrator: Command Prompt
19788 ResourceManager
C:\WINDOWS\system32>ls
'ls' is not recognized as an internal or external command,
operable program or batch file.
C:\WINDOWS\system32>hdfs dfs -ls /
Found 3 items
drwxr-xr-x - Aarya supergroup 0 2020-09-28 15:03 /nkk
drwxr-xr-x - Aarya supergroup 0 2020-10-19 16:24 /spark
drwxr-xr-x - Aarya supergroup 0 2020-09-28 15:23 /user
C:\WINDOWS\system32>hdfs dfs -mkdir niknew
C:\WINDOWS\system32>hdfs dfs -ls /
Found 3 items
drwxr-xr-x - Aarya supergroup 0 2020-09-28 15:03 /nkk
drwxr-xr-x - Aarya supergroup 0 2020-10-19 16:24 /spark
drwxr-xr-x - Aarya supergroup 0 2020-09-28 15:23 /user
C:\WINDOWS\system32>
Type here to search


Administrator: Command Prompt
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons
C:\WINDOWS\system32>jps
15520 NameNode
7584 NodeManager
17144 DataNode
1564 Jps
19788 ResourceManager
C:\WINDOWS\system32>ls
'ls' is not recognized as an internal or external command,
operable program or batch file.
C:\WINDOWS\system32>hdfs dfs -ls /
Found 3 items
drwxr-xr-x - Aarya supergroup 0 2020-09-28 15:03 /nkk
drwxr-xr-x - Aarya supergroup 0 2020-10-19 16:24 /spark
drwxr-xr-x - Aarya supergroup 0 2020-09-28 15:23 /user
C:\WINDOWS\system32>
Type here to search
10:27 PM
02-Dec-20


Administrator: Command Prompt
Microsoft Windows [Version 10.0.19042.630]
(c) 2020 Microsoft Corporation. All rights reserved.
C:\WINDOWS\system32>jps
19120 Jps
C:\WINDOWS\system32>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons
C:\WINDOWS\system32>jps
15520 NameNode
7584 NodeManager
17144 DataNode
1564 Jps
19788 ResourceManager
C:\WINDOWS\system32>
Type here to search
10:27 PM
02-Dec-20

AIM-MAPREDUCE WORD COUNT USING COMMANDS

STEPS-

Create an input directory in HDFS.

hadoop fs -mkdir /input_dir

Copy the input text file named input_file.txt in the input directory (input_dir)of HDFS.

hadoop fs -put C:/input_file.txt /input_dir

Verify input_file.txt available in HDFS input directory (input_dir).

hadoop fs -ls /input_dir/

Input_file

Verify content of the copied file.

hadoop dfs -cat /input_dir/input_file.txt

Content

Run MapReduceClient.jar and also provide input and out directories.

hadoop jar C:/MapReduceClient.jar wordcount /input_dir /output_dir

Success

Verify content for generated output file.

hadoop dfs -cat /output_dir/*

out

Some Other usefull commands

To leave Safe mode

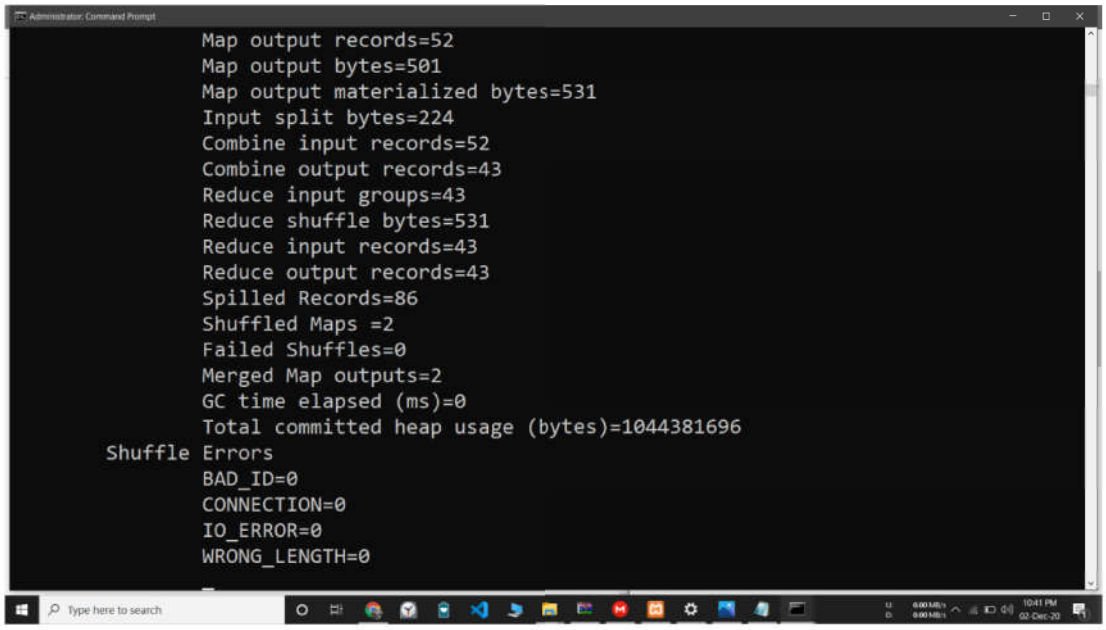
hadoop dfsadmin –safemode leave

To Delete file from HDFS directory

hadoop fs -rm -r /iutput_dir/input_file.txt

To Delete directory from HDFS directory

hadoop fs -rm -r /iutput_dir


Administrator: Command Prompt
Map output records=52
Map output bytes=501
Map output materialized bytes=531
Input split bytes=224
Combine input records=52
Combine output records=43
Reduce input groups=43
Reduce shuffle bytes=531
Reduce input records=43
Reduce output records=43
Spilled Records=86
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=0
Total committed heap usage (bytes)=1044381696
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
Type here to search
10:41 PM
02-Dec-20

AIM-INSTALLATION OF SPARK

STEPS

Install Apache Spark on Windows

Installing Apache Spark on Windows 10 may seem complicated to novice users, but this simple tutorial will have you up and running. If you already have Java 8 and Python 3 installed, you can skip the first two steps.

Step 1: Install Java 8

Apache Spark requires Java 8. You can check to see if Java is installed using the command prompt.

Open the command line by clicking Start > type cmd > click Command Prompt.

Type the following command in the command prompt:

java -version

If Java is installed, it will respond with the following output:

Windows CLI output for Java version.

Your version may be different. The second digit is the Java version – in this case, Java 8.

If you don't have Java installed:

1. Open a browser window, and navigate to https://java.com/en/download/.

Java download page in a browser

2. Click the Java Download button and save the file to a location of your choice.

3. Once the download finishes double-click the file to install Java.

Note: At the time this article was written, the latest Java version is 1.8.0_251. Installing a later version will still work. This process only needs the Java Runtime Environment (JRE) – the full Development Kit (JDK) is not required. The download link to JDK is https://www.oracle.com/java/technologies/javase-downloads.html.

Step 2: Install Python

1. To install the Python package manager, navigate to https://www.python.org/ in your web browser.

2. Mouse over the Download menu option and click Python 3.8.3. 3.8.3 is the latest version at the time of writing the article.

3. Once the download finishes, run the file.

Python download page for version 3.8.3

4. Near the bottom of the first setup dialog box, check off Add Python 3.8 to PATH. Leave the other box checked.

5. Next, click Customize installation.

Python wizard 3.8.3, step to add Python to PATH

6. You can leave all boxes checked at this step, or you can uncheck the options you do not want.

7. Click Next.

8. Select the box Install for all users and leave other boxes as they are.

9. Under Customize install location, click Browse and navigate to the C drive. Add a new folder and name it Python.

10. Select that folder and click OK.

Python installation, advanced options step

11. Click Install, and let the installation complete.

12. When the installation completes, click the Disable path length limit option at the bottom and then click Close.

13. If you have a command prompt open, restart it. Verify the installation by checking the version of Python:

python –-version

The output should print Python 3.8.3.

Note: For detailed instructions on how to install Python 3 on Windows or how to troubleshoot potential issues, refer to our Install Python 3 on Windows guide.

Step 3: Download Apache Spark

1. Open a browser and navigate to https://spark.apache.org/downloads.html.

2. Under the Download Apache Spark heading, there are two drop-down menus. Use the current non-preview version.

In our case, in Choose a Spark release drop-down menu select 2.4.5 (Feb 05 2020).

In the second drop-down Choose a package type, leave the selection Pre-built for Apache Hadoop 2.7.

3. Click the spark-2.4.5-bin-hadoop2.7.tgz link.

Apache Spark download page.

4. A page with a list of mirrors loads where you can see different servers to download from. Pick any from the list and save the file to your Downloads folder.

Step 4: Verify Spark Software File

1. Verify the integrity of your download by checking the checksum of the file. This ensures you are working with unaltered, uncorrupted software.

2. Navigate back to the Spark Download page and open the Checksum link, preferably in a new tab.

3. Next, open a command line and enter the following command:

certutil -hashfile c:\users\username\Downloads\spark-2.4.5-bin-hadoop2.7.tgz SHA512

4. Change the username to your username. The system displays a long alphanumeric code, along with the message Certutil: -hashfile completed successfully.

Checksum output for the Spark installation file.

5. Compare the code to the one you opened in a new browser tab. If they match, your download file is uncorrupted.

Step 5: Install Apache Spark

Installing Apache Spark involves extracting the downloaded file to the desired location.

1. Create a new folder named Spark in the root of your C: drive. From a command line, enter the following:

```
cd \

mkdir Spark
```

2. In Explorer, locate the Spark file you downloaded.

3. Right-click the file and extract it to C:\Spark using the tool you have on your system (e.g., 7-Zip).

4. Now, your C:\Spark folder has a new folder spark-2.4.5-bin-hadoop2.7 with the necessary files inside.

Step 6: Add winutils.exe File

Download the winutils.exe file for the underlying Hadoop version for the Spark installation you downloaded.

1. Navigate to this URL https://github.com/cdarlint/winutils and inside the bin folder, locate winutils.exe, and click it.

Winutils download page

2. Find the Download button on the right side to download the file.

3. Now, create new folders Hadoop and bin on C: using Windows Explorer or the Command Prompt.

4. Copy the winutils.exe file from the Downloads folder to C:\hadoop\bin.

Step 7: Configure Environment Variables

This step adds the Spark and Hadoop locations to your system PATH. It allows you to run the Spark shell directly from a command prompt window.

1. Click Start and type environment.

2. Select the result labeled Edit the system environment variables.

3. A System Properties dialog box appears. In the lower-right corner, click Environment Variables and then click New in the next window.

Add new environment variable in Windows.

4. For Variable Name type SPARK_HOME.

5. For Variable Value type C:\Spark\spark-2.4.5-bin-hadoop2.7 and click OK. If you changed the folder path, use that one instead.

Adding Spark home variable path in Windows.

6. In the top box, click the Path entry, then click Edit. Be careful with editing the system path. Avoid deleting any entries already on the list.

Edit the path variable to add Spark home.

7. You should see a box with entries on the left. On the right, click New.

8. The system highlights a new line. Enter the path to the Spark folder C:\Spark\spark-2.4.5-bin-hadoop2.7\bin. We recommend using %SPARK_HOME%\bin to avoid possible issues with the path.

Adding the Spark home to the path Windows variable.

9. Repeat this process for Hadoop and Java.

For Hadoop, the variable name is HADOOP_HOME and for the value use the path of the folder you created earlier: C:\hadoop. Add C:\hadoop\bin to the Path variable field, but we recommend using %HADOOP_HOME%\bin.

For Java, the variable name is JAVA_HOME and for the value use the path to your Java JDK directory (in our case it's C:\Program Files\Java\jdk1.8.0_251).

10. Click OK to close all open windows.

Note: Star by restarting the Command Prompt to apply changes. If that doesn't work, you will need to reboot the system.

Step 8: Launch Spark

1. Open a new command-prompt window using the right-click and Run as administrator:

2. To start Spark, enter:

C:\Spark\spark-2.4.5-bin-hadoop2.7\bin\spark-shell

If you set the environment path correctly, you can type spark-shell to launch Spark.

3. The system should display several lines indicating the status of the application. You may get a Java pop-up. Select Allow access to continue.

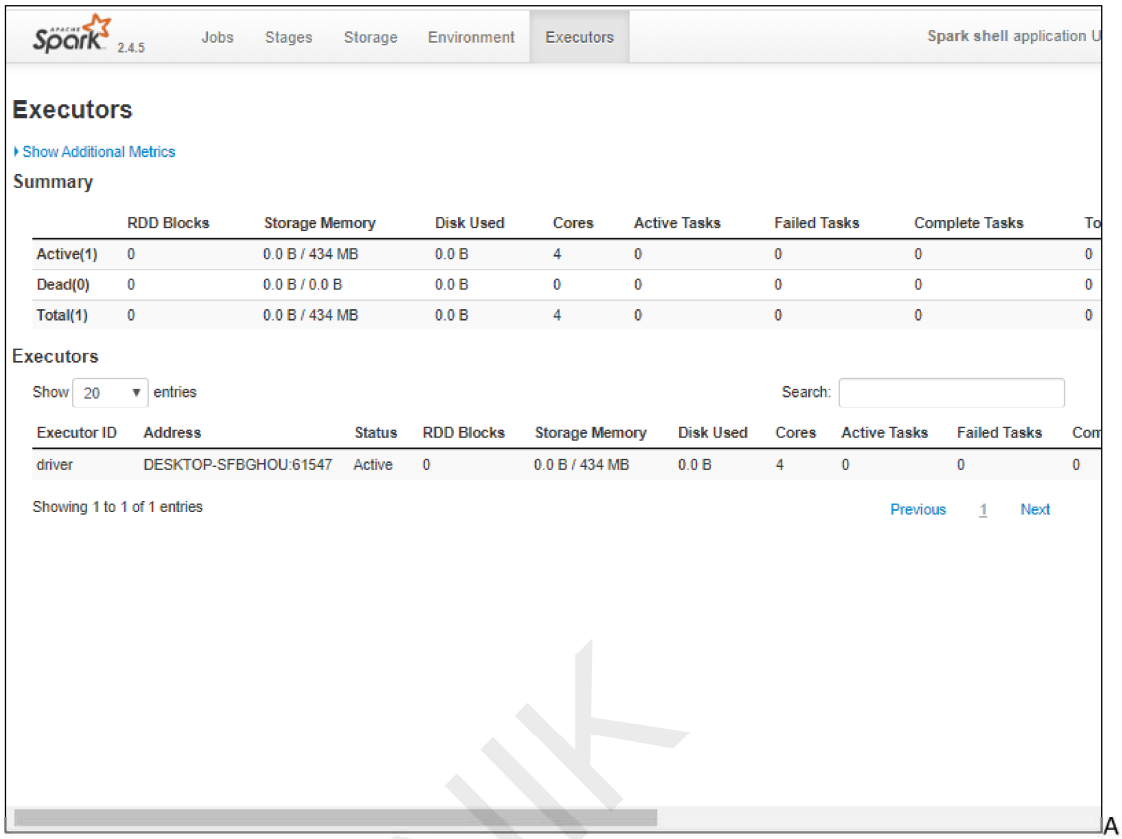
Finally, the Spark logo appears, and the prompt displays the Scala shell.

Scala shell after launching apacheSpark in windows

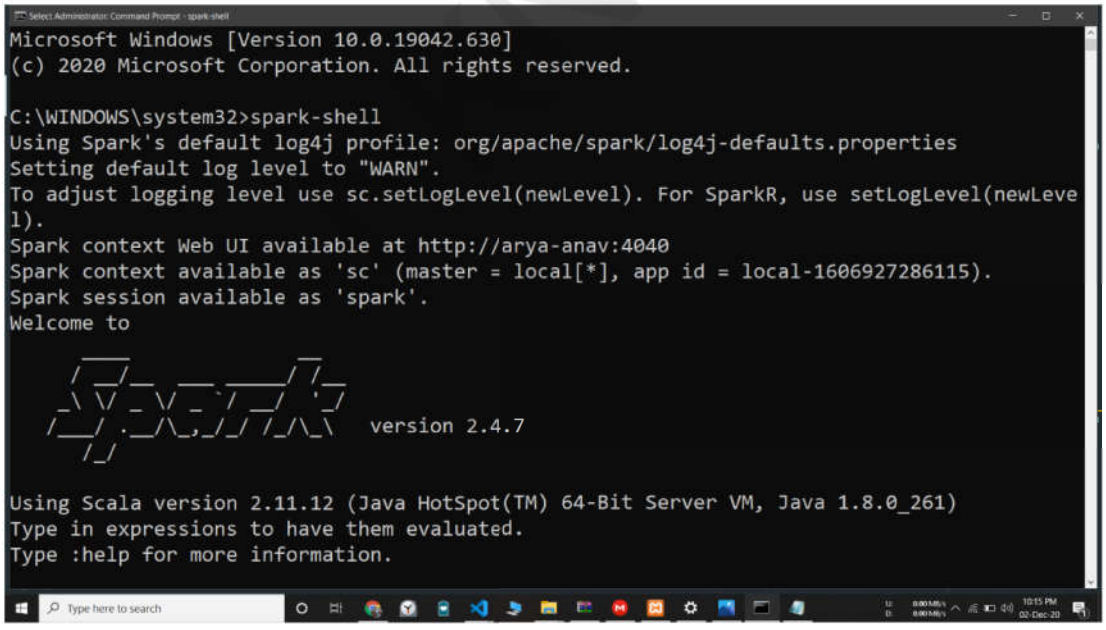
4., Open a web browser and navigate to http://localhost:4040/.

5. You can replace localhost with the name of your system.

6. You should see an Apache Spark shell Web UI. The example below shows the Executors page.


Spark 2.4.5
Jobs
Stages
Storage
Environment
Executors
Spark shell application U
Executors
Show Additional Metrics
Summary
RDD Blocks
Storage Memory
Disk Used
Cores
Active Tasks
Failed Tasks
Complete Tasks
To
Active(1) 0 0.0 B / 434 MB 0.0 B 4 0 0 0 0
Dead(0) 0 0.0 B / 0.0 B 0.0 B 0 0 0 0 0
Total(1) 0 0.0 B / 434 MB 0.0 B 4 0 0 0 0
Executors
Show 20 entries
Search:
Executor ID
Address
Status
RDD Blocks
Storage Memory
Disk Used
Cores
Active Tasks
Failed Tasks
Com
driver DESKTOP-SFBGHOU:61547 Active 0 0.0 B / 434 MB 0.0 B 4 0 0 0
Showing 1 to 1 of 1 entries
Previous
1
Next

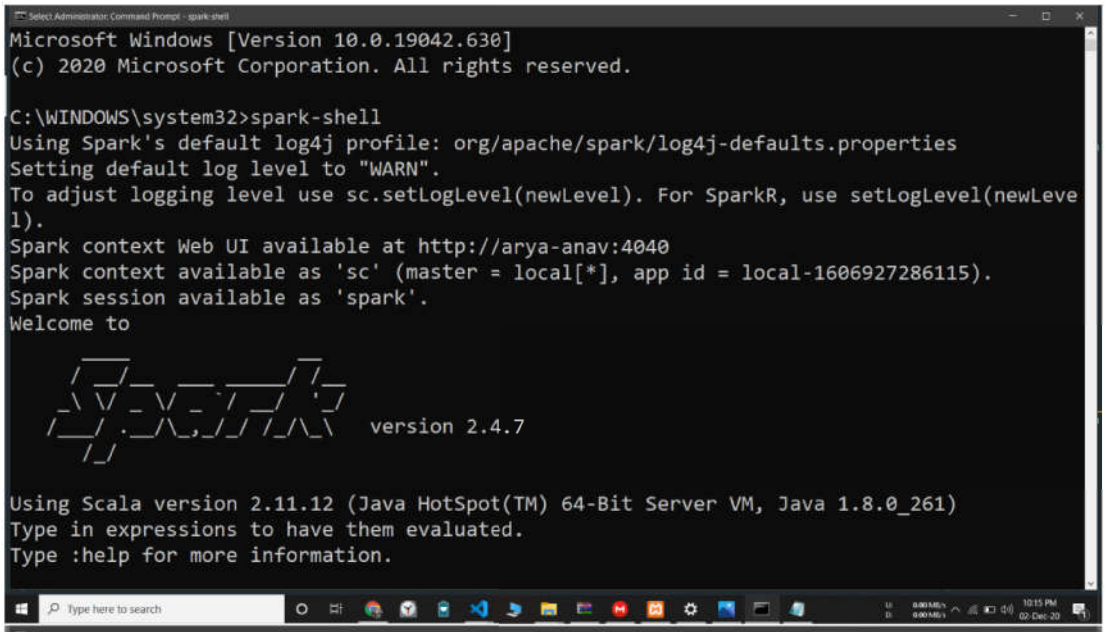
A

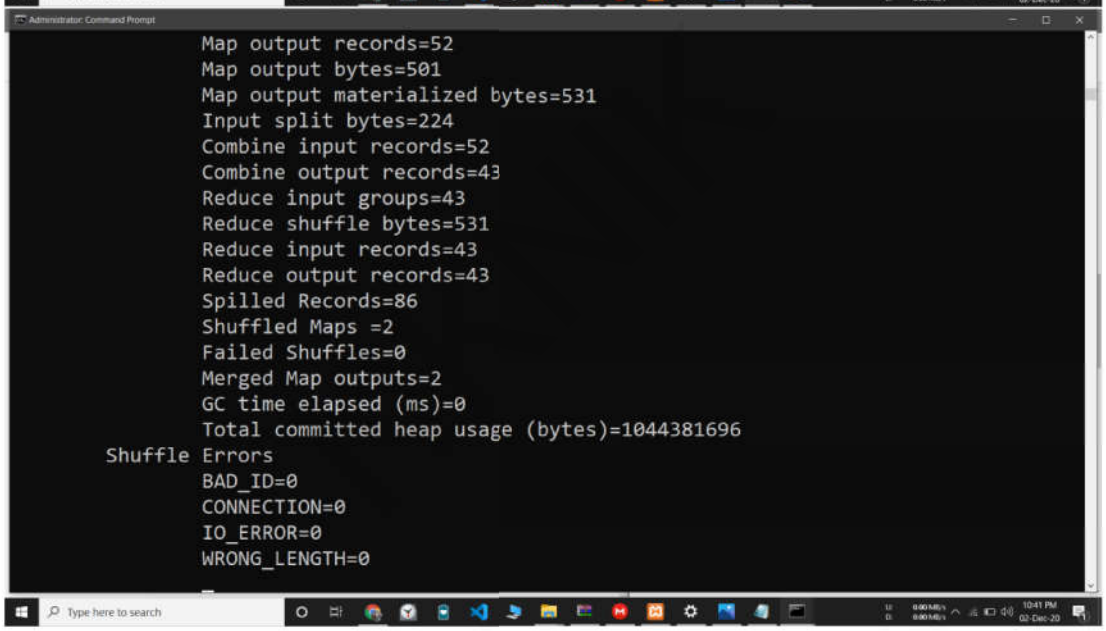

Microsoft Windows [Version 10.0.19042.630]
(c) 2020 Microsoft Corporation. All rights reserved.
C:\WINDOWS\system32>spark-shell
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLeve
l).
Spark context Web UI available at http://arya-anav:4040
Spark context available as 'sc' (master = local[*], app id = local-1606927286115).
Spark session available as 'spark'.
Welcome to
version 2.4.7
Using Scala version 2.11.12 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_261)
Type in expressions to have them evaluated.
Type :help for more information.
Type here to search

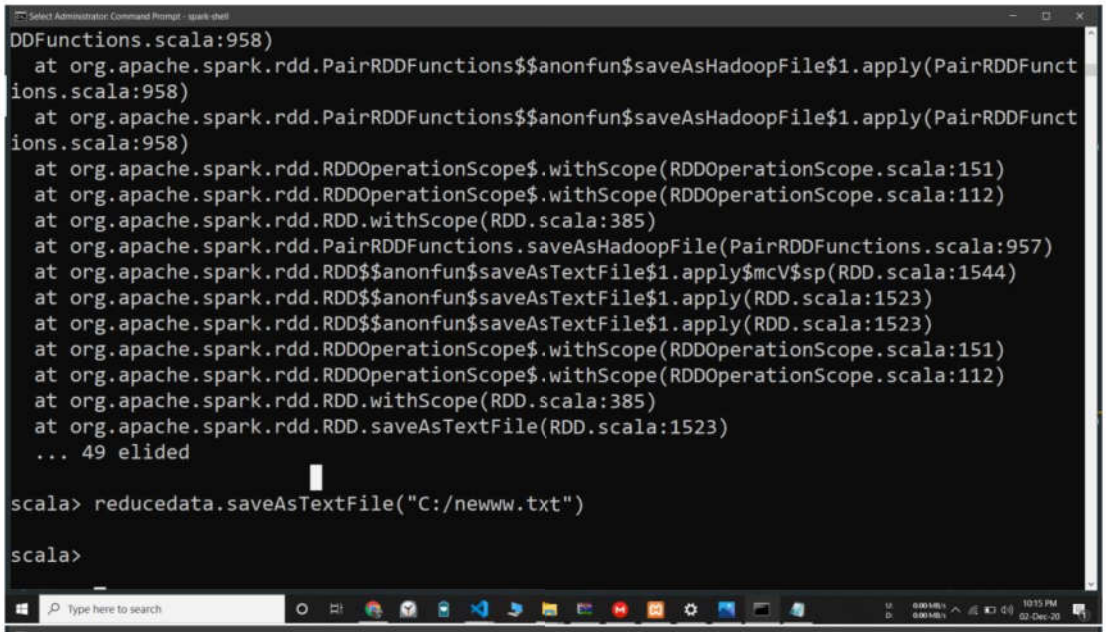
AIM-WORD COUNT USING SPARK

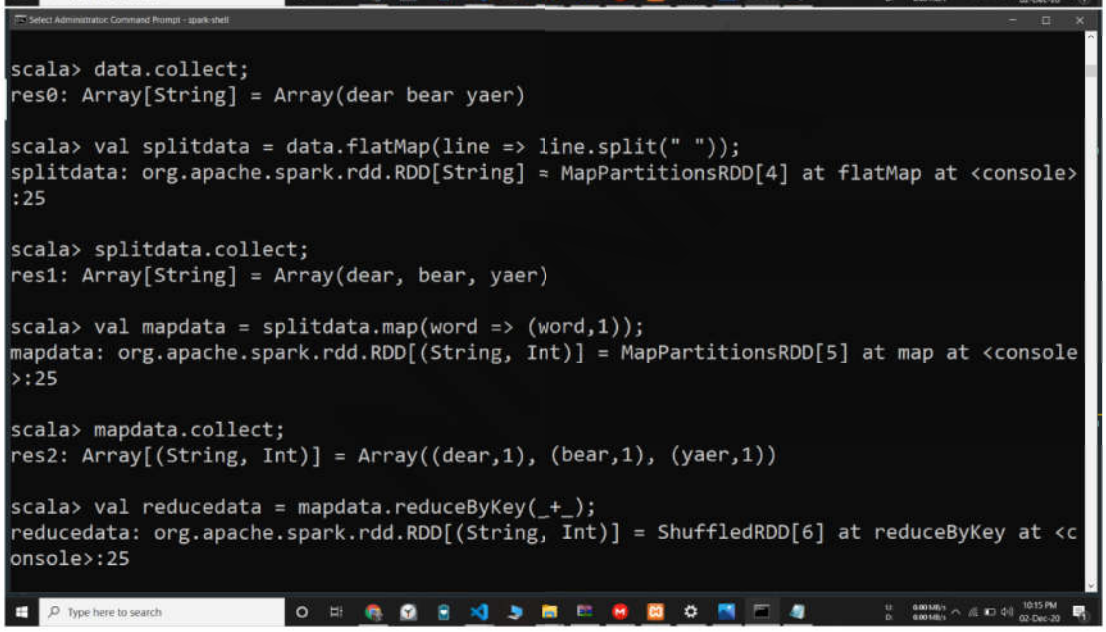
STEPS-

```
val data=sc.textFile("N:/wordcount.txt")

data.collect;

val splitdata = data.flatMap(line => line.split(" "));

splitdata.collect;

val mapdata = splitdata.map(word => (word,1));

 mapdata.collect;

 val reducedata = mapdata.reduceByKey(_+_);

 reducedata.collect;

 reducedata.saveAsTextFile("N:/scala/word_scala")
```


Select Administrator: Command Prompt - spark-shell
Microsoft Windows [Version 10.0.19042.630]
(c) 2020 Microsoft Corporation. All rights reserved.
C:\WINDOWS\system32>spark-shell
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Spark context Web UI available at http://arya-anav:4040
Spark context available as 'sc' (master = local[*], app id = local-1606927286115).
Spark session available as 'spark'.
Welcome to
version 2.4.7
Using Scala version 2.11.12 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_261)
Type in expressions to have them evaluated.
Type :help for more information.
Type here to search
10:15 PM
02-Dec-20


Administrator: Command Prompt
Map output records=52
Map output bytes=501
Map output materialized bytes=531
Input split bytes=224
Combine input records=52
Combine output records=43
Reduce input groups=43
Reduce shuffle bytes=531
Reduce input records=43
Reduce output records=43
Spilled Records=86
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=0
Total committed heap usage (bytes)=1044381696
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
Type here to search
10:41 PM
02-Dec-20


Select Administrator: Command Prompt - spark-shell
DDFunctions.scala:958)
at org.apache.spark.rdd.PairRDDFunctions$$anonfun$saveAsHadoopFile$1.apply(PairRDDFunct
ions.scala:958)
at org.apache.spark.rdd.PairRDDFunctions$$anonfun$saveAsHadoopFile$1.apply(PairRDDFunct
ions.scala:958)
at org.apache.spark.rdd.RDDOperationScope$.withScope(RDDOperationScope.scala:151)
at org.apache.spark.rdd.RDDOperationScope$.withScope(RDDOperationScope.scala:112)
at org.apache.spark.rdd.RDD.withScope(RDD.scala:385)
at org.apache.spark.rdd.PairRDDFunctions.saveAsHadoopFile(PairRDDFunctions.scala:957)
at org.apache.spark.rdd.RDD$$anonfun$saveAsTextFile$1.applymcVsp(RDD.scala:1544)
at org.apache.spark.rdd.RDD$$anonfun$saveAsTextFile$1.apply(RDD.scala:1523)
at org.apache.spark.rdd.RDD$$anonfun$saveAsTextFile$1.apply(RDD.scala:1523)
at org.apache.spark.rdd.RDDOperationScope$.withScope(RDDOperationScope.scala:151)
at org.apache.spark.rdd.RDDOperationScope$.withScope(RDDOperationScope.scala:112)
at org.apache.spark.rdd.RDD.withScope(RDD.scala:385)
at org.apache.spark.rdd.RDD.saveAsTextFile(RDD.scala:1523)
... 49 elided
scala> reducedata.saveAsTextFile("C:/newww.txt")
scala>
Type here to search
10:15 PM
02-Dec-20


Select Administrator: Command Prompt - spark-shell
scala> data.collect;
res0: Array[String] = Array(dear bear yaer)
scala> val splitdata = data.flatMap(line => line.split(" "));
splitdata: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[4] at flatMap at <console>
:25
scala> splitdata.collect;
res1: Array[String] = Array(dear, bear, yaer)
scala> val mapdata = splitdata.map(word => (word,1));
mapdata: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[5] at map at <console
>:25
scala> mapdata.collect;
res2: Array[(String, Int)] = Array((dear,1), (bear,1), (yaer,1))
scala> val reducedata = mapdata.reduceByKey(_+_);
reducedata: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[6] at reduceByKey at <c
onsole>:25
Type here to search
10:15 PM
02-Dec-20

AIM-INSTALLATION OF HBASE

STEPS-

HBase Installation Steps:

Step 1:

Unzip the downloaded Hbase and place it in some common path, say C:/Document/hbase-2.2.5

Step 2:

Create a folders as shown below inside root folder for HBase data and zookeeper

-> C:/Document/hbase-2.2.5/hbase

-> C:/Document/hbase-2.2.5/zookeeper

Step 3:

Open C:/Document/hbase-2.2.5/bin/hbase.cmd in notepad++. Search for below given lines and remove %HEAP_SETTINGS% from that line as dictated in the video embedded with this blog

set java_arguments=%HEAP_SETTINGS% %HBASE_OPTS% -classpath "%CLASSPATH%" %CLASS% %hbase-command-arguments%

Step 4:

Open C:/Document/hbase-2.2.5/conf/hbase-env.cmd n notepad++. Add the below lines to the file after the comment session as shown in the YT video given with this tutorial.

set JAVA_HOME=%JAVA_HOME%

set HBASE_CLASSPATH=%HBASE_HOME%\lib\client-facing-thirdparty*

set HBASE_HEAPSIZE=8000

set HBASE_OPTS="-XX:+UseConcMarkSweepGC" "-Djava.net.preferIPv4Stack=true"

set SERVER_GC_OPTS="-verbose:gc" "-XX:+PrintGCDetails" "-XX:+PrintGCDateStamps" %HBASE_GC_OPTS%

set HBASE_USE_GC_LOGFILE=true

set HBASE_JMX_BASE="-Dcom.sun.management.jmxremote.ssl=false" "-Dcom.sun.management.jmxremote.authenticate=false"

set HBASE_MASTER_OPTS=%HBASE_JMX_BASE% "-Dcom.sun.management.jmxremote.port=10101"

set HBASE_REGIONSERVER_OPTS=%HBASE_JMX_BASE% "-Dcom.sun.management.jmxremote.port=10102"

set HBASE_THRIFT_OPTS=%HBASE_JMX_BASE% "-Dcom.sun.management.jmxremote.port=10103"

set HBASE_ZOOKEEPER_OPTS=%HBASE_JMX_BASE% -Dcom.sun.management.jmxremote.port=10104"

set HBASE_REGIONSERVERS=%HBASE_HOME%\conf\regionservers

set HBASE_LOG_DIR=%HBASE_HOME%\logs

set HBASE_IDENT_STRING=%USERNAME%

set HBASE_MANAGES_ZK=true

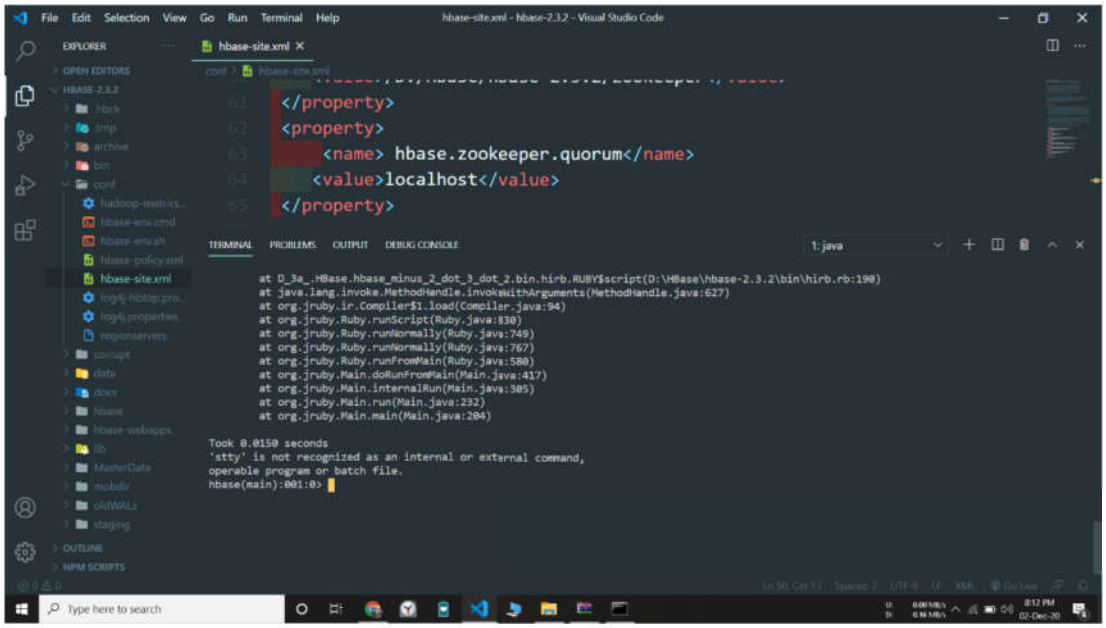
Step 5:

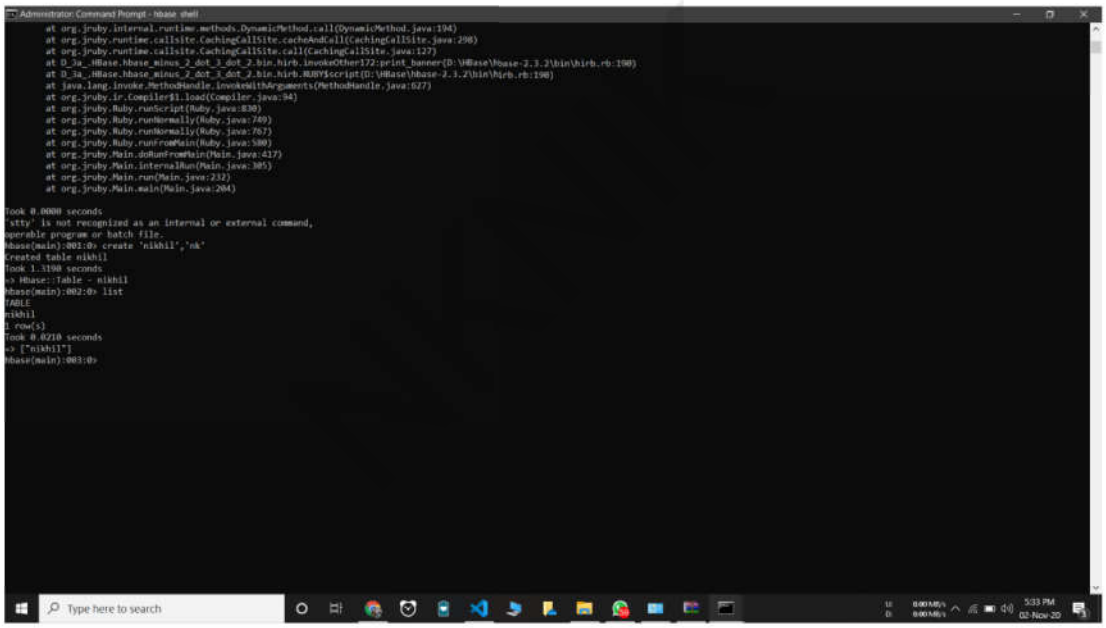
Open C:/Document/hbase-2.2.5/conf/hbase-site.xml notepad++. Add the below lines inside <configuration> tag. Refer YT video given in this tutorial.

```
<property>
  <name>hbase.rootdir</name>
  <value>file:///C:/Documents/hbase-2.2.5/hbase</value>
</property>
<property>
  <name>hbase.zookeeper.property.dataDir</name>
  <value>/C:/Documents/hbase-2.2.5/zookeeper</value>
</property>
<property>
   <name> hbase.zookeeper.quorum</name>
  <value>localhost</value>
</property>
```

Step 6:

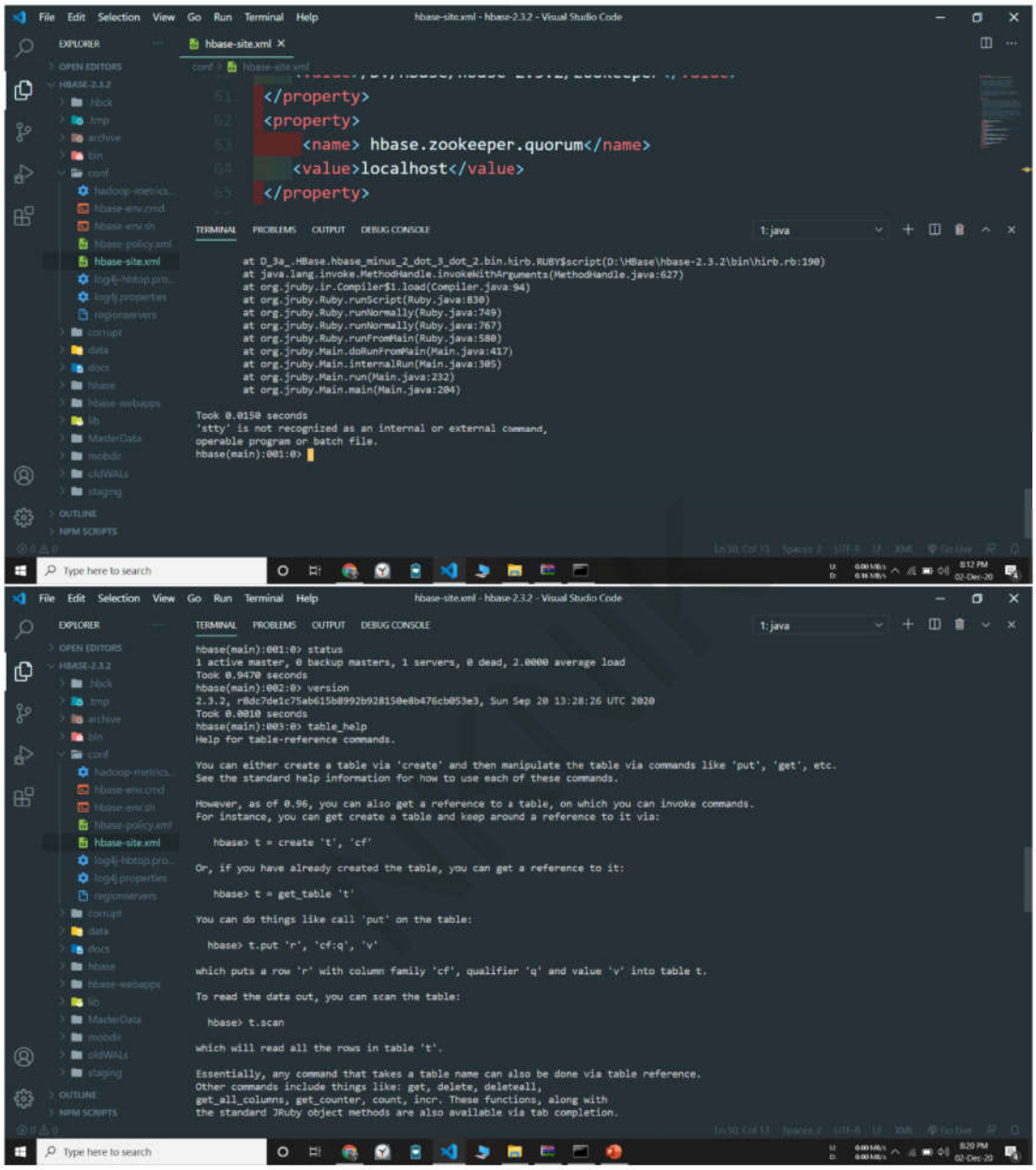
Setup the Environment variable for HBASE_HOME and add bin to the path variable as shown in the below image.

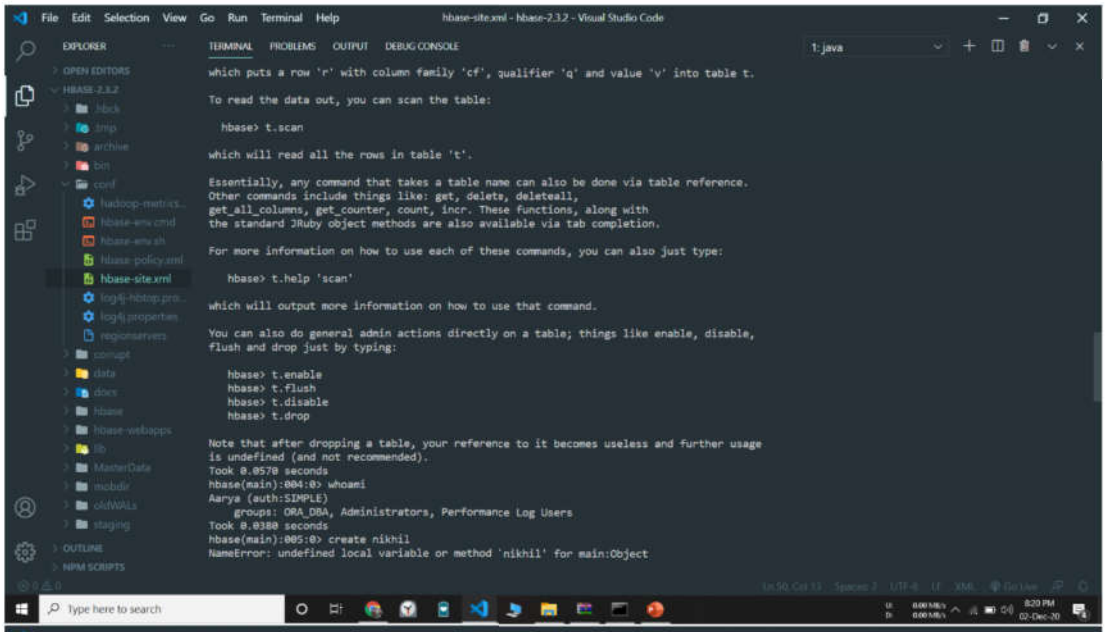

File Edit Selection View Go Run Terminal Help
hbase-site.xml - hbase-2.3.2 - Visual Studio Code
EXPLORER
hbase-site.xml
conf > hbase-site.xml
</property>
<property>
<name> hbase.zookeeper.quorum</name>
<value>localhost</value>
</property>
TERMINAL PROBLEMS OUTPUT DEBUG CONSOLE
1: java
at D_3a_.HBase.hbase_minus_2_dot_3_dot_2.bin.hirb.RUBY$script(D:\HBase\hbase-2.3.2\bin\hirb.rb:190)
at java.lang.invoke.MethodHandle.invokeWithArguments(MethodHandle.java:627)
at org.jruby.ir.Compiler$1.load(Compiler.java:94)
at org.jruby.Ruby.runScript(Ruby.java:830)
at org.jruby.Ruby.runNormally(Ruby.java:749)
at org.jruby.Ruby.runNormally(Ruby.java:767)
at org.jruby.Ruby.runFromMain(Ruby.java:580)
at org.jruby.Main.doRunFromMain(Main.java:417)
at org.jruby.Main.internalRun(Main.java:305)
at org.jruby.Main.run(Main.java:232)
at org.jruby.Main.main(Main.java:204)
Took 0.0150 seconds
'stty' is not recognized as an internal or external command,
operable program or batch file.
hbase(main):001:0>
Type here to search
8:12 PM
02-Dec-20

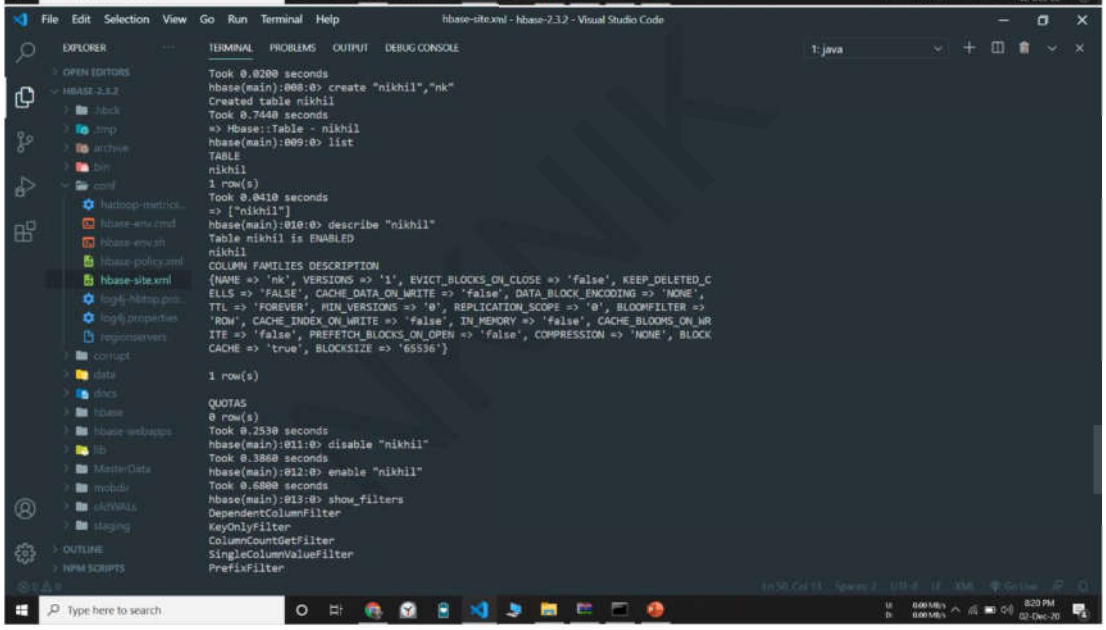

Administrator: Command Prompt - hbase shell
at org.jruby.internal.runtime.methods.DynamicMethod.call(DynamicMethod.java:194)
at org.jruby.runtime.callsite.CachingCallSite.cacheAndCall(CachingCallSite.java:298)
at org.jruby.runtime.callsite.CachingCallSite.call(CachingCallSite.java:127)
at D_3a_.HBase.hbase_minus_2_dot_3_dot_2.bin.hirb.invokeOther172:print_banner(D:\HBase\hbase-2.3.2\bin\hirb.rb:190)
at D_3a_.HBase.hbase_minus_2_dot_3_dot_2.bin.hirb.RUBY$script(D:\HBase\hbase-2.3.2\bin\hirb.rb:190)
at java.lang.invoke.MethodHandle.invokeWithArguments(MethodHandle.java:627)
at org.jruby.ir.Compiler$1.load(Compiler.java:94)
at org.jruby.Ruby.runScript(Ruby.java:830)
at org.jruby.Ruby.runNormally(Ruby.java:749)
at org.jruby.Ruby.runNormally(Ruby.java:767)
at org.jruby.Ruby.runFromMain(Ruby.java:580)
at org.jruby.Main.doRunFromMain(Main.java:417)
at org.jruby.Main.internalRun(Main.java:305)
at org.jruby.Main.run(Main.java:232)
at org.jruby.Main.main(Main.java:204)
Took 0.0000 seconds
'stty' is not recognized as an internal or external command,
operable program or batch file.
hbase(main):003:0> create 'nikhil','nk'
Created table nikhil
Took 1.3190 seconds
=> Hbase::Table - nikhil
hbase(main):002:0> list
TABLE
nikhil
1 row(s)
Took 0.0210 seconds
=> ["nikhil"]
hbase(main):003:0>
Type here to search
5:33 PM
02-Nov-20

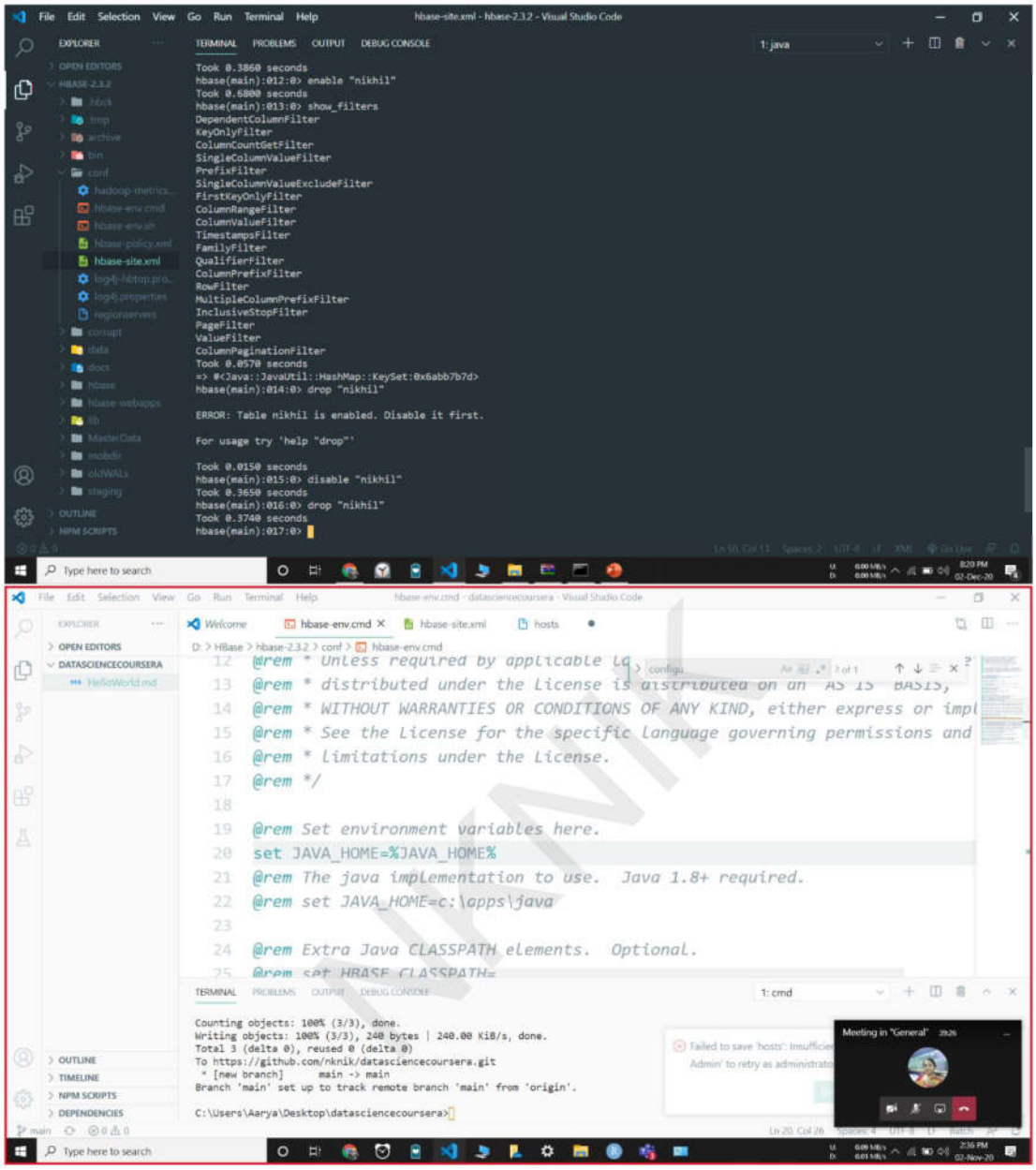
AIM-DDL AND DML COMMANDS IN HBASE

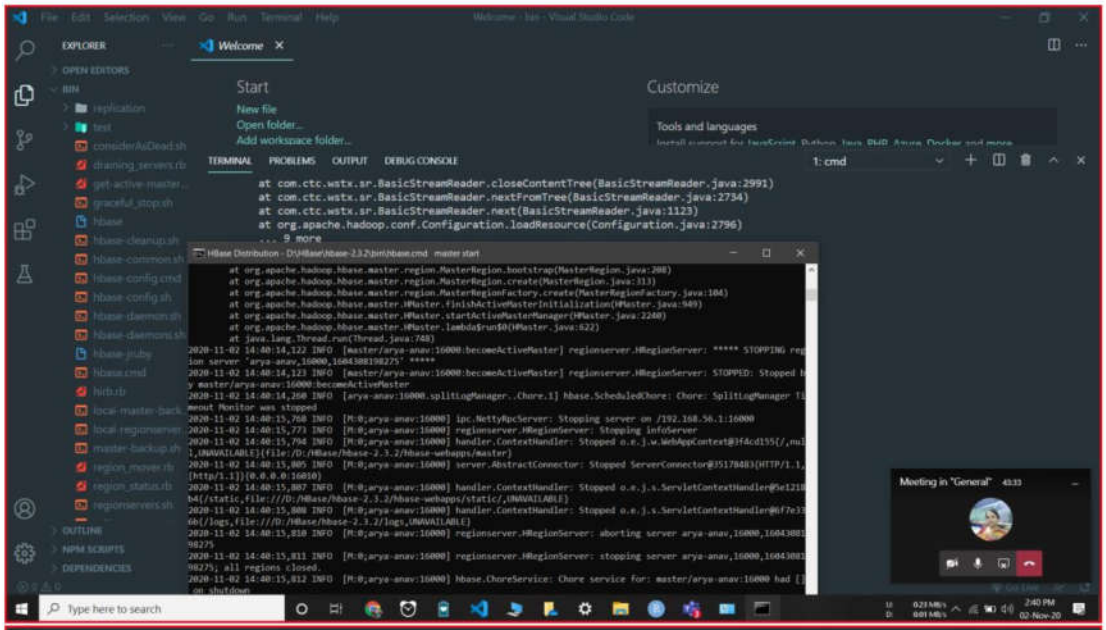
STEPS-

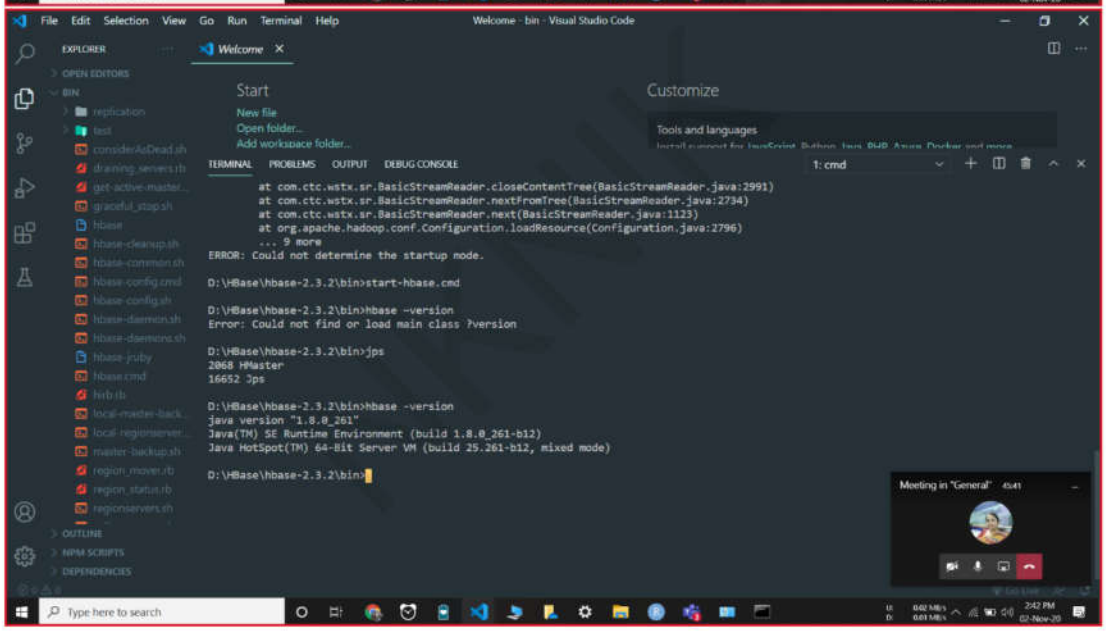


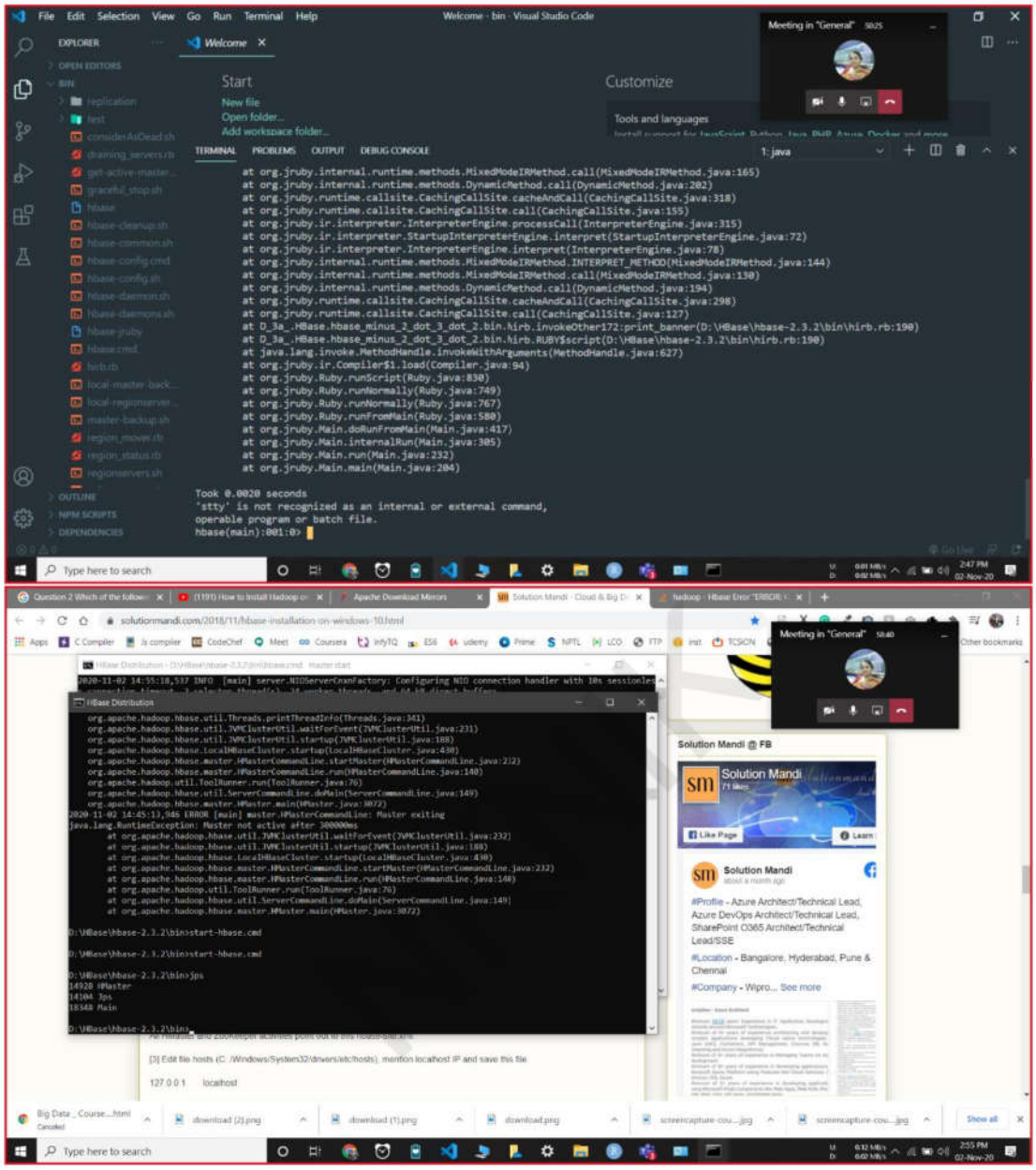

File Edit Selection View Go Run Terminal Help
hbase-site.xml - hbase-2.3.2 - Visual Studio Code
EXPLORER
TERMINAL PROBLEMS OUTPUT DEBUG CONSOLE
1: java
which puts a row 'r' with column family 'cf', qualifier 'q' and value 'v' into table t.
To read the data out, you can scan the table:
hbase> t.scan
which will read all the rows in table 't'.
Essentially, any command that takes a table name can also be done via table reference.
Other commands include things like: get, delete, deleteall,
get_all_columns, get_counter, count, incr. These functions, along with
the standard JRuby object methods are also available via tab completion.
For more information on how to use each of these commands, you can also just type:
hbase> t.help 'scan'
which will output more information on how to use that command.
You can also do general admin actions directly on a table; things like enable, disable,
flush and drop just by typing:
hbase> t.enable
hbase> t.flush
hbase> t.disable
hbase> t.drop
Note that after dropping a table, your reference to it becomes useless and further usage
is undefined (and not recommended).
Took 0.0570 seconds
hbase(main):004:0> whoami
Aarya (auth:SIMPLE)
groups: ORA_DBA, Administrators, Performance Log Users
Took 0.0380 seconds
hbase(main):005:0> create nikhil
NameError: undefined local variable or method `nikhil' for main:Object
Type here to search
8:20 PM
02-Dec-20

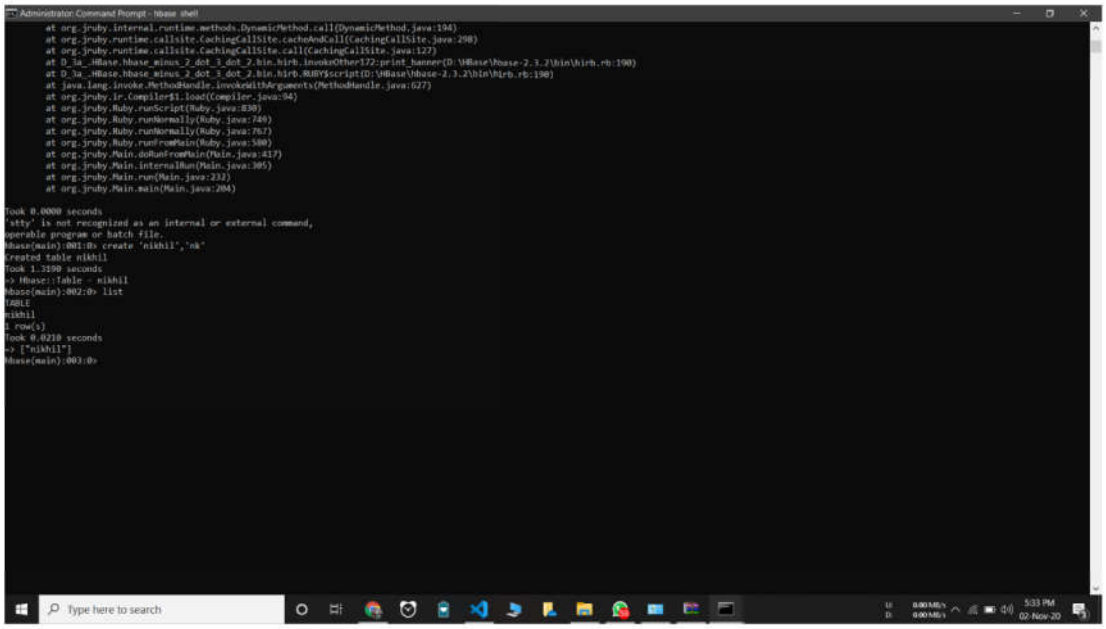

File Edit Selection View Go Run Terminal Help
hbase-site.xml - hbase-2.3.2 - Visual Studio Code
EXPLORER
TERMINAL PROBLEMS OUTPUT DEBUG CONSOLE
1: java
Took 0.0200 seconds
hbase(main):008:0> create "nikhil","nk"
Created table nikhil
Took 0.7440 seconds
=> Hbase::Table - nikhil
hbase(main):009:0> list
TABLE
nikhil
1 row(s)
Took 0.0410 seconds
=> ["nikhil"]
hbase(main):010:0> describe "nikhil"
Table nikhil is ENABLED
nikhil
COLUMN FAMILIES DESCRIPTION
{NAME => 'nk', VERSIONS => '1', EVICT_BLOCKS_ON_CLOSE => 'false', KEEP_DELETED_C
ELLS => 'FALSE', CACHE_DATA_ON_WRITE => 'false', DATA_BLOCK_ENCODING => 'NONE',
TTL => 'FOREVER', MIN_VERSIONS => '0', REPLICATION_SCOPE => '0', BLOOMFILTER =>
'ROW', CACHE_INDEX_ON_WRITE => 'false', IN_MEMORY => 'false', CACHE_BLOOMS_ON_WR
ITE => 'false', PREFETCH_BLOCKS_ON_OPEN => 'false', COMPRESSION => 'NONE', BLOCK
CACHE => 'true', BLOCKSIZE => '65536'}
1 row(s)
QUOTAS
0 row(s)
Took 0.2530 seconds
hbase(main):011:0> disable "nikhil"
Took 0.3860 seconds
hbase(main):012:0> enable "nikhil"
Took 0.6880 seconds
hbase(main):013:0> show_filters
DependentColumnFilter
KeyOnlyFilter
ColumnCountGetFilter
SingleColumnValueFilter
PrefixFilter
Type here to search
8:20 PM
02-Dec-20


File Edit Selection View Go Run Terminal Help
hbase-site.xml - hbase-2.3.2 - Visual Studio Code
EXPLORER
TERMINAL PROBLEMS OUTPUT DEBUG CONSOLE
1: java
Took 0.3860 seconds
hbase(main):012:0> enable "nikhil"
Took 0.6800 seconds
hbase(main):013:0> show_filters
DependentColumnFilter
KeyOnlyFilter
ColumnCountGetFilter
SingleColumnValueFilter
PrefixFilter
SingleColumnValueExcludeFilter
FirstKeyOnlyFilter
ColumnRangeFilter
ColumnValueFilter
TimestampsFilter
FamilyFilter
QualifierFilter
ColumnPrefixFilter
RowFilter
MultipleColumnPrefixFilter
InclusiveStopFilter
PageFilter
ValueFilter
ColumnPaginationFilter
Took 0.0570 seconds
=> #<Java::JavaUtil::HashMap::KeySet:0x6abb7b7d>
hbase(main):014:0> drop "nikhil"
ERROR: Table nikhil is enabled. Disable it first.
For usage try 'help "drop"'
Took 0.0150 seconds
hbase(main):015:0> disable "nikhil"
Took 0.3650 seconds
hbase(main):016:0> drop "nikhil"
Took 0.3740 seconds
hbase(main):017:0>
Type here to search
8:20 PM
02-Dec-20
hbase-env.cmd - datasciencecoursera - Visual Studio Code
Welcome
hbase-env.cmd
hbase-site.xml
hosts
D: > HBase > hbase-2.3.2 > conf > hbase-env.cmd
OPEN EDITORS
DATASCIENCECOURSERA
HelloWorld.md
@rem * Unless required by applicable la
@rem * distributed under the License is distributed on an "AS IS" BASIS,
@rem * WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or impl
@rem * See the License for the specific language governing permissions and
@rem * limitations under the License.
@rem */
@rem Set environment variables here.
set JAVA_HOME=%JAVA_HOME%
@rem The java implementation to use. Java 1.8+ required.
@rem set JAVA_HOME=c:\apps\java
@rem Extra Java CLASSPATH elements. Optional.
@rem set HBASE_CLASSPATH=
TERMINAL PROBLEMS OUTPUT DEBUG CONSOLE
1: cmd
Counting objects: 100% (3/3), done.
Writing objects: 100% (3/3), 240 bytes | 240.00 KiB/s, done.
Total 3 (delta 0), reused 0 (delta 0)
To https://github.com/nknik/datasciencecoursera.git
* [new branch] main -> main
Branch 'main' set up to track remote branch 'main' from 'origin'.
C:\Users\Aarya\Desktop\datasciencecoursera>
Meeting in "General"
OUTLINE
TIMELINE
NPM SCRIPTS
DEPENDENCIES
Type here to search
2:36 PM
02-Nov-20


Welcome - bin - Visual Studio Code
Welcome
Start
New file
Open folder...
Add workspace folder...
Customize
Tools and languages
TERMINAL
PROBLEMS
OUTPUT
DEBUG CONSOLE
1: cmd
at com.ctc.wstx.sr.BasicStreamReader.closeContentTree(BasicStreamReader.java:2991)
at com.ctc.wstx.sr.BasicStreamReader.nextFromTree(BasicStreamReader.java:2734)
at com.ctc.wstx.sr.BasicStreamReader.next(BasicStreamReader.java:1123)
at org.apache.hadoop.conf.Configuration.loadResource(Configuration.java:2796)
... 9 more
HBase Distribution - D:\HBase\hbase-2.3.2\bin\hbase.cmd master start
Meeting in "General"
Type here to search

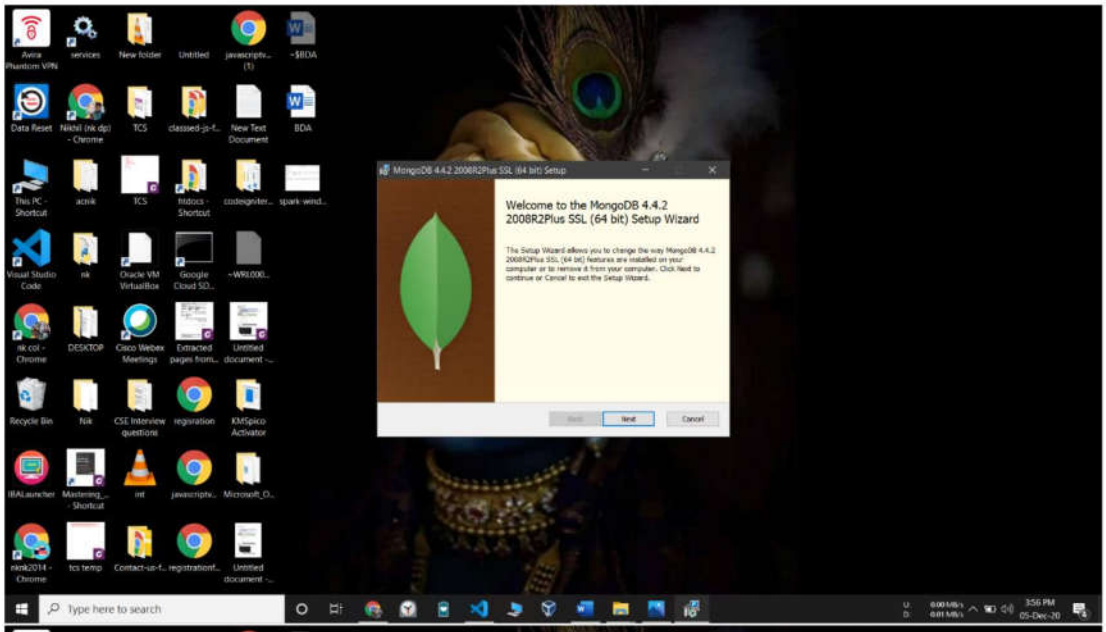

Welcome - bin - Visual Studio Code
Welcome
Start
New file
Open folder...
Add workspace folder...
Customize
Tools and languages
TERMINAL
PROBLEMS
OUTPUT
DEBUG CONSOLE
1: cmd
at com.ctc.wstx.sr.BasicStreamReader.closeContentTree(BasicStreamReader.java:2991)
at com.ctc.wstx.sr.BasicStreamReader.nextFromTree(BasicStreamReader.java:2734)
at com.ctc.wstx.sr.BasicStreamReader.next(BasicStreamReader.java:1123)
at org.apache.hadoop.conf.Configuration.loadResource(Configuration.java:2796)
... 9 more
ERROR: Could not determine the startup mode.
D:\HBase\hbase-2.3.2\bin>start-hbase.cmd
D:\HBase\hbase-2.3.2\bin>hbase -version
Error: Could not find or load main class ?version
D:\HBase\hbase-2.3.2\bin>jps
2868 HMaster
16652 Jps
D:\HBase\hbase-2.3.2\bin>hbase -version
java version "1.8.0_261"
Java(TM) SE Runtime Environment (build 1.8.0_261-b12)
Java HotSpot(TM) 64-Bit Server VM (build 25.261-b12, mixed mode)
D:\HBase\hbase-2.3.2\bin>
Meeting in "General"
Type here to search

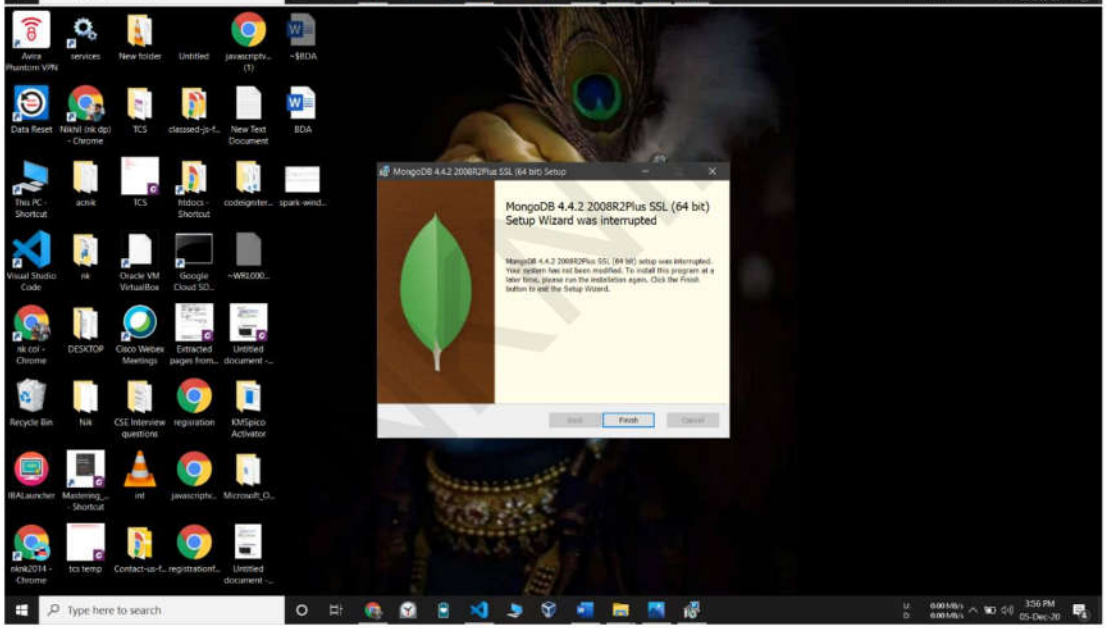

File Edit Selection View Go Run Terminal Help
Welcome - bin - Visual Studio Code
EXPLORER
Welcome
Start
New file
Open folder...
Add workspace folder...
Customize
Tools and languages
TERMINAL PROBLEMS OUTPUT DEBUG CONSOLE
1: java
at org.jruby.internal.runtime.methods.MixedModeIRMethod.call(MixedModeIRMethod.java:165)
at org.jruby.internal.runtime.methods.DynamicMethod.call(DynamicMethod.java:202)
at org.jruby.runtime.callsite.CachingCallSite.cacheAndCall(CachingCallSite.java:318)
at org.jruby.runtime.callsite.CachingCallSite.call(CachingCallSite.java:159)
at org.jruby.ir.interpreter.InterpreterEngine.processCall(InterpreterEngine.java:315)
at org.jruby.ir.interpreter.StartupInterpreterEngine.interpret(StartupInterpreterEngine.java:72)
at org.jruby.ir.interpreter.InterpreterEngine.interpret(InterpreterEngine.java:78)
at org.jruby.internal.runtime.methods.MixedModeIRMethod.INTERPRET_METHOD(MixedModeIRMethod.java:144)
at org.jruby.internal.runtime.methods.MixedModeIRMethod.call(MixedModeIRMethod.java:130)
at org.jruby.internal.runtime.methods.DynamicMethod.call(DynamicMethod.java:194)
at org.jruby.runtime.callsite.CachingCallSite.cacheAndCall(CachingCallSite.java:298)
at org.jruby.runtime.callsite.CachingCallSite.call(CachingCallSite.java:127)
at D_3a_.HBase.hbase_minus_2_dot_3_dot_2.bin.hirb.invokeOther172:print_banner(D:\HBase\hbase-2.3.2\bin\hirb.rb:190)
at D_3a_.HBase.hbase_minus_2_dot_3_dot_2.bin.hirb.RUBY$script(D:\HBase\hbase-2.3.2\bin\hirb.rb:190)
at java.lang.invoke.MethodHandle.invokeWithArguments(MethodHandle.java:627)
at org.jruby.ir.Compiler$1.load(Compiler.java:94)
at org.jruby.Ruby.runScript(Ruby.java:830)
at org.jruby.Ruby.runNormally(Ruby.java:749)
at org.jruby.Ruby.runNormally(Ruby.java:767)
at org.jruby.Ruby.runFromMain(Ruby.java:580)
at org.jruby.Main.doRunFromMain(Main.java:417)
at org.jruby.Main.internalRun(Main.java:305)
at org.jruby.Main.run(Main.java:232)
at org.jruby.Main.main(Main.java:204)
Took 0.0020 seconds
'stty' is not recognized as an internal or external command,
operable program or batch file.
hbase(main):001:0>
Meeting in "General"
Type here to search
HBase Distribution
org.apache.hadoop.hbase.util.Threads.printThreadInfo(Threads.java:341)
org.apache.hadoop.hbase.util.JVMClusterUtil.waitForEvent(JVMClusterUtil.java:231)
org.apache.hadoop.hbase.util.JVMClusterUtil.startup(JVMClusterUtil.java:188)
org.apache.hadoop.hbase.LocalHBaseCluster.startup(LocalHBaseCluster.java:430)
org.apache.hadoop.hbase.master.HMasterCommandLine.startMaster(HMasterCommandLine.java:232)
org.apache.hadoop.hbase.master.HMasterCommandLine.run(HMasterCommandLine.java:140)
org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:76)
org.apache.hadoop.hbase.util.ServerCommandLine.doMain(ServerCommandLine.java:149)
org.apache.hadoop.hbase.master.HMaster.main(HMaster.java:3072)
2020-11-02 14:45:13,945 ERROR [main] master.HMasterCommandLine: Master exiting
java.lang.RuntimeException: Master not active after 30000ms
D:\HBase\hbase-2.3.2\bin>start-hbase.cmd
D:\HBase\hbase-2.3.2\bin>jps
14928 HMaster
14104 Jps
18548 Main
[3] Edit file hosts (C: /Windows/System32/drivers/etc/hosts), mention localhost IP and save this file
127.0.0.1 localhost
Solution Mandi @ FB
Solution Mandi
Like Page
#Profile - Azure Architect/Technical Lead, Azure DevOps Architect/Technical Lead, SharePoint O365 Architect/Technical Lead/SSE
#Location - Bangalore, Hyderabad, Pune & Chennai
#Company - Wipro... See more
Meeting in "General"
Type here to search

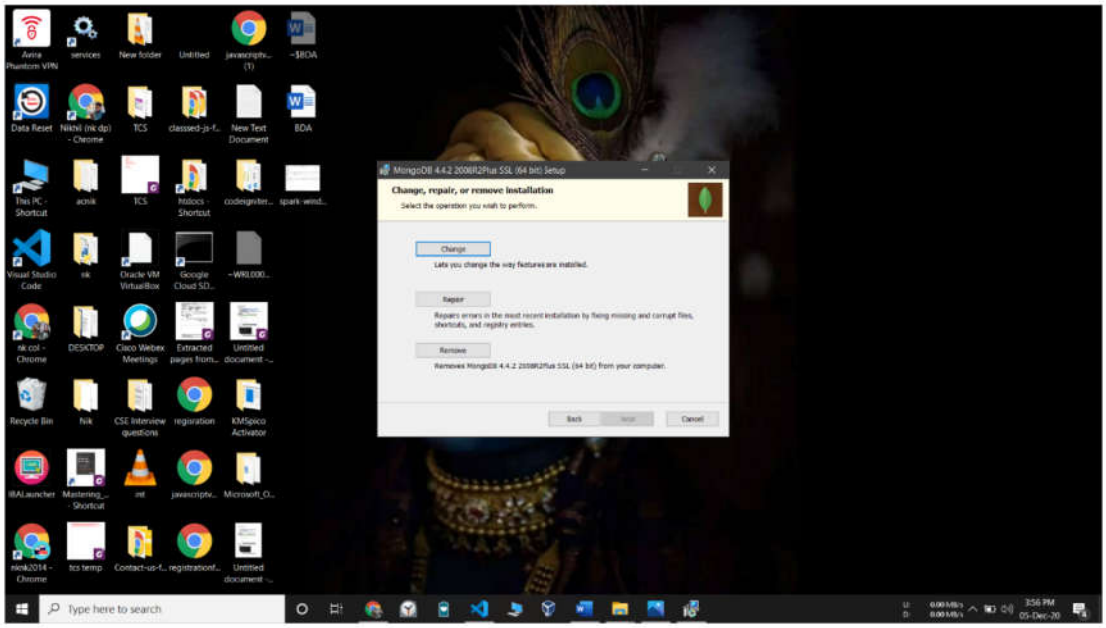

Administrator: Command Prompt - hbase shell
at org.jruby.internal.runtime.methods.DynamicMethod.call(DynamicMethod.java:194)
at org.jruby.runtime.callsite.CachingCallSite.cacheAndCall(CachingCallSite.java:298)
at org.jruby.runtime.callsite.CachingCallSite.call(CachingCallSite.java:127)
at D_3a__HBase.hbase_minus_2_dot_3_dot_2.bin.hirb.invokeOther172:print_banner(D:\HBase\hbase-2.3.2\bin\hirb.rb:190)
at D_3a__HBase.hbase_minus_2_dot_3_dot_2.bin.hirb.RUBY$script(D:\HBase\hbase-2.3.2\bin\hirb.rb:190)
at java.lang.invoke.MethodHandle.invokeWithArguments(MethodHandle.java:627)
at org.jruby.ir.Compiler$1.load(Compiler.java:94)
at org.jruby.Ruby.runScript(Ruby.java:830)
at org.jruby.Ruby.runNormally(Ruby.java:749)
at org.jruby.Ruby.runNormally(Ruby.java:767)
at org.jruby.Ruby.runFromMain(Ruby.java:580)
at org.jruby.Main.doRunFromMain(Main.java:417)
at org.jruby.Main.internalRun(Main.java:305)
at org.jruby.Main.run(Main.java:232)
at org.jruby.Main.main(Main.java:204)
Took 0.0000 seconds
'stty' is not recognized as an internal or external command,
operable program or batch file.
hbase(main):001:0> create 'nikhil','nk'
Created table nikhil
Took 1.3190 seconds
=> Hbase::Table - nikhil
hbase(main):002:0> list
TABLE
nikhil
1 row(s)
Took 0.0210 seconds
=> ["nikhil"]
hbase(main):003:0>
Type here to search
5:33 PM
02-Nov-20

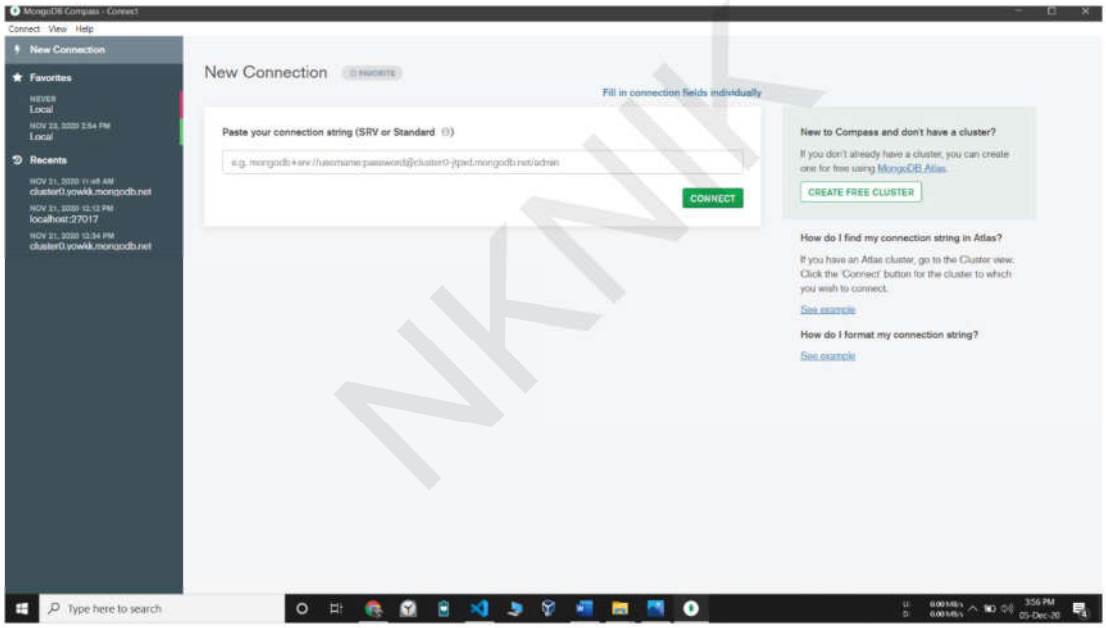
AIM-INSTALLATION OF MONGODB

STEPS-


MongoDB 4.4.2 2008R2Plus SSL (64 bit) Setup
Welcome to the MongoDB 4.4.2 2008R2Plus SSL (64 bit) Setup Wizard
The Setup Wizard allows you to change the way MongoDB 4.4.2 2008R2Plus SSL (64 bit) features are installed on your computer or to remove it from your computer. Click Next to continue or Cancel to exit the Setup Wizard.
Next
Cancel
Type here to search
3:56 PM
05-Dec-20


MongoDB 4.4.2 2008R2Plus SSL (64 bit) Setup
MongoDB 4.4.2 2008R2Plus SSL (64 bit) Setup Wizard was interrupted
MongoDB 4.4.2 2008R2Plus SSL (64 bit) setup was interrupted. Your system has not been modified. To install this program at a later time, please run the installation again. Click the Finish button to exit the Setup Wizard.
Finish
Cancel
Type here to search
3:56 PM
05-Dec-20


MongoDB 4.4.2 2008R2Plus SSL (64 bit) Setup
Change, repair, or remove installation
Select the operation you wish to perform.
Change
Lets you change the way features are installed.
Repair
Repairs errors in the most recent installation by fixing missing and corrupt files, shortcuts, and registry entries.
Remove
Removes MongoDB 4.4.2 2008R2Plus SSL (64 bit) from your computer.
Back
Next
Cancel
Type here to search


MongoDB Compass - Connect
Connect View Help
New Connection
Favorites
Local
Local
Recents
cluster0.yowkk.mongodb.net
localhost:27017
cluster0.yowkk.mongodb.net
New Connection
Fill in connection fields individually
Paste your connection string (SRV or Standard)
e.g. mongodb+srv://username:password@cluster0-jqad.mongodb.net/admin
CONNECT
New to Compass and don't have a cluster?
If you don't already have a cluster, you can create one for free using MongoDB Atlas.
CREATE FREE CLUSTER
How do I find my connection string in Atlas?
If you have an Atlas cluster, go to the Cluster view. Click the 'Connect' button for the cluster to which you wish to connect.
See example
How do I format my connection string?
See example
Type here to search

AIM-BASIC CRUD OPERATIONS IN MONGODB

STEPS-

C:\Program Files\MongoDB\Server\4.4\bin>mongo

MongoDB shell version v4.4.2

connecting to: mongodb://127.0.0.1:27017/?compressors=disabled&gssapiServiceName=mongodb

Implicit session: session { "id" : UUID("8b9a07c9-a207-4666-910f-9df6d9cee29c") }

MongoDB server version: 4.4.2

The server generated these startup warnings when booting:

2020-11-23T13:37:12.232+05:30: Access control is not enabled for the database. Read and write access to data and configuration is unrestricted

---

Enable MongoDB's free cloud-based monitoring service, which will then receive and display metrics about your deployment (disk utilization, CPU, operation statistics, etc).

The monitoring data will be available on a MongoDB website with a unique URL accessible to you and anyone you share the URL with. MongoDB may use this information to make product improvements and to suggest MongoDB products and deployment options to you.

To enable free monitoring, run the following command: db.enableFreeMonitoring()

To permanently disable this reminder, run the following command: db.disableFreeMonitoring()

> use b2

switched to db b2

> show db

uncaught exception: Error: don't know how to show [db] :

shellHelper.show@src/mongo/shell/utils.js:1191:11

shellHelper@src/mongo/shell/utils.js:819:15

@(shellhelp2):1:1

> db

```
b2
> show dbs
admin    0.000GB
config   0.000GB
local    0.000GB
travell  0.001GB
> db
b2
> db.dropDatabase()
{ "ok" : 1 }
> use b2
switched to db b2
> db.createCollection(name,option)
uncaught exception: ReferenceError: name is not defined :
@(shell):1:1
> db.createCollection("try1",{capped: true, autoIndexId: true, size : 6142800, max:1000})
{
        "note" : "The autoIndexId option is deprecated and will be removed in a future release",
        "ok" : 1
}
> show collections
try1
> db.try1.drop()
true
> db.createCollection("try1",{capped: true, autoIndexId: true, size : 6142800, max:1000})
{
        "note" : "The autoIndexId option is deprecated and will be removed in a future release",
        "ok" : 1
}
> db.try1.insert ( { “name”: “nk”})
uncaught exception: SyntaxError: illegal character :
```

@(shell):1:19

> db.try1.insert({ “name”: “nk”})

uncaught exception: SyntaxError: illegal character :

@(shell):1:18

> db.try1.insert({“name”: “nk”})

uncaught exception: SyntaxError: illegal character :

@(shell):1:16

> db.try1.insert({“name”:“nk”})

uncaught exception: SyntaxError: illegal character :

@(shell):1:16

> db.try1.insert({'name':'nk'})

WriteResult({ "nInserted" : 1 })

> show collections

try1

> db.try1.find()

{ "_id" : ObjectId("5fbb7d5ab84d48fd31d20c22"), "name" : "nk" }

> db.try1.find().pritty()

uncaught exception: TypeError: db.try1.find(...).pritty is not a function :

@(shell):1:1

> db.try1.find().pretty()

{ db.try1.insert({ _id : ObjectId ("012345678901234567890123") , name : "NK", title : "Mongodb", roll_no : 56, email_id : ['jj@gmail.com','kk@gmail.com','j@gmail.com']})nsert({ _id : ObjectId ("012345678901234567890123") , name : "NK", title : "Mongodb", roll_no : 56, email_id : ['jj@gmail.com','kk@gmail.com','j@gmaiWriteResult({ "nInserted" : 1 })

> db.try1.find().pretty()

{ "_id" : ObjectId("5fbb7d5ab84d48fd31d20c22"), "name" : "nk" }

{

"_id" : ObjectId("012345678901234567890123"),

"name" : "NK",

"title" : "Mongodb",

"roll_no" : 56,

"email_id" : [

```
            "jj@gmail.com",
            "kk@gmail.com",
            "j@gmail.com"
        ]
}
> db.try1.insertMany([ { name: "Amaan", roll_no :75, fees: 1299}, {name: "vaishnavi", roll_no : 81,fees: 12345}])
{
        "acknowledged" : true,
        "insertedIds" : [
                ObjectId("5fbb7e25b84d48fd31d20c23"),
                ObjectId("5fbb7e25b84d48fd31d20c24")
        ]
}
> db.try1.find({"name":"Sony"}).pretty()
> db.try1.find({"name":"nk"}).pretty()
{ "_id" : ObjectId("5fbb7d5ab84d48fd31d20c22"), "name" : "nk" }
> db.try1.find().pretty()
{ "_id" : ObjectId("5fbb7d5ab84d48fd31d20c22"), "name" : "nk" }
{
        "_id" : ObjectId("012345678901234567890123"),
        "name" : "NK",
        "title" : "Mongodb",
        "roll_no" : 56,
        "email_id" : [
                "jj@gmail.com",
                "kk@gmail.com",
                "j@gmail.com"
        ]
}
{
        "_id" : ObjectId("5fbb7e25b84d48fd31d20c23"),
```

```
        "name" : "Amaan",
        "roll_no" : 75,
        "fees" : 1299
}
{
        "_id" : ObjectId("5fbb7e25b84d48fd31d20c24"),
        "name" : "vaishnavi",
        "roll_no" : 81,
        "fees" : 12345
}
> db.collection_name.find({ $and :[{key:value},{key:value}]})
uncaught exception: ReferenceError: value is not defined :
@(shell):1:35
> db.collection_name.find({ $and :[{key:value},{key:value}]})
uncaught exception: ReferenceError: value is not defined :
@(shell):1:35
> db.try1.find( { $and : [{ "name":"Amaan"},{"fees":1299}]}).pretty()
{
        "_id" : ObjectId("5fbb7e25b84d48fd31d20c23"),
        "name" : "Amaan",
        "roll_no" : 75,
        "fees" : 1299
}
> db.try1.find({"fees": { $not: {$gt: 10000}}}).pretty()
{ "_id" : ObjectId("5fbb7d5ab84d48fd31d20c22"), "name" : "nk" }
{
        "_id" : ObjectId("012345678901234567890123"),
        "name" : "NK",
        "title" : "Mongodb",
        "roll_no" : 56,
        "email_id" : [
```

```
            "jj@gmail.com",
            "kk@gmail.com",
            "j@gmail.com"
        ]
}
{
        "_id" : ObjectId("5fbb7e25b84d48fd31d20c23"),
        "name" : "Amaan",
        "roll_no" : 75,
        "fees" : 1299
}
> db.try.1update({'fees':1299},{$set:{'fees':123456789}})
uncaught exception: SyntaxError: identifier starts immediately after numeric literal :
@(shell):1:6
> db.try1.update({'fees':1299},{$set:{'fees':123456789}})
WriteResult({ "nMatched" : 1, "nUpserted" : 0, "nModified" : 1 })
> db.try1.updateOne({Name:'JK'},{$set: {Exam:'PT2'}})
{ "acknowledged" : true, "matchedCount" : 0, "modifiedCount" : 0 }
> db.try1.updateMany({fees:{$gt: 10000}}, {$set: {fees: 12345}})
{ "acknowledged" : true, "matchedCount" : 2, "modifiedCount" : 1 }
> db.try1.remove({ Name : ‘Vaibhav’})
uncaught exception: SyntaxError: illegal character :
@(shell):1:24
> db.try1.remove({ Name : 'Vaibhav'})
WriteResult({
        "nRemoved" : 0,
        "writeError" : {
            "code" : 20,
            "errmsg" : "cannot remove from a capped collection: b2.try1"
        }
})
```