Greenest Part

AIM: To calculate the areas in bangalore with area covered in greenery covered over 75%

Preprocessing:

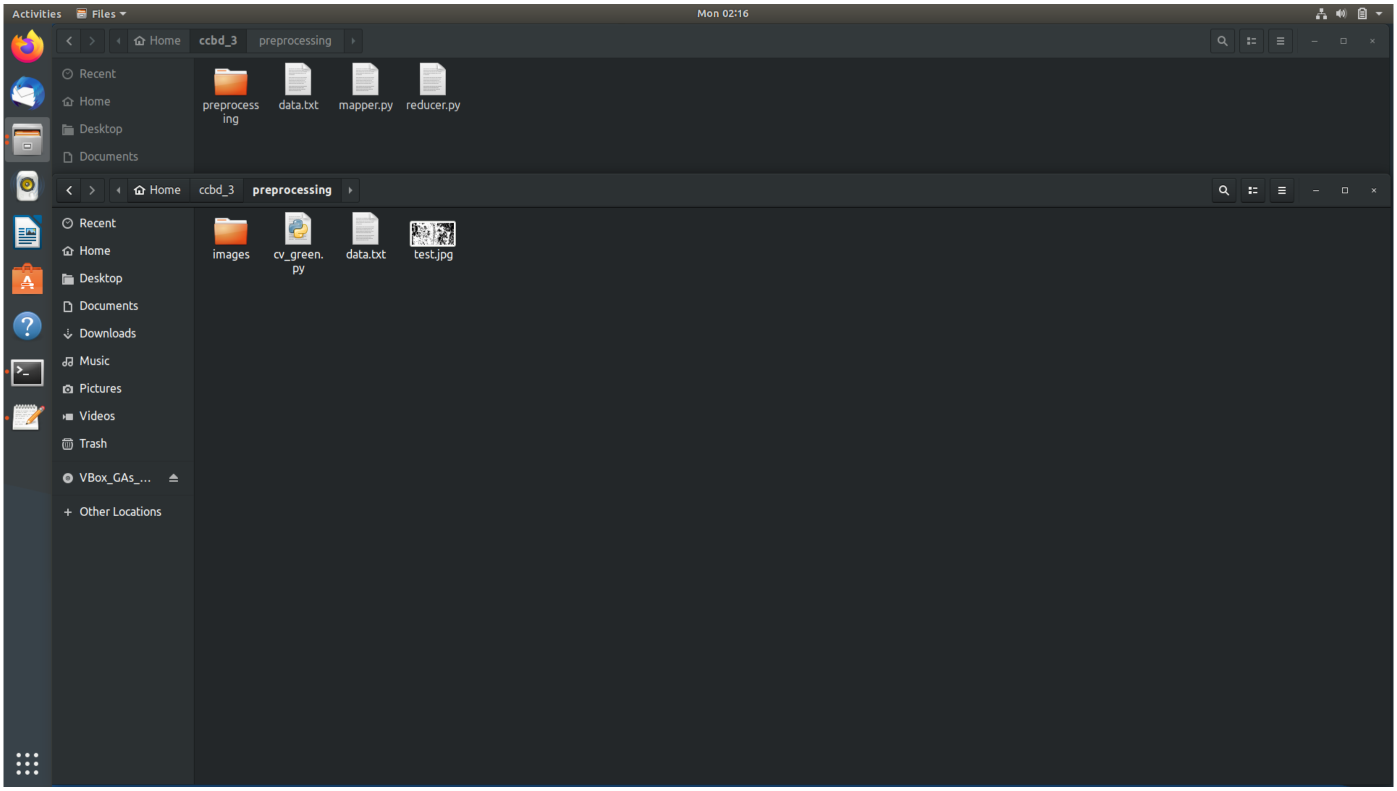
- Install opencv using the command:

  ```
  sudo apt install python3-opencv
  ```

  Install numpy using the command:

  ```
  pip3 install numpy
  ```

- Login into ***username*** using your password.
- Clone this repository and copy it into a folder of your choice:
  - In this case: `ccbd_3`



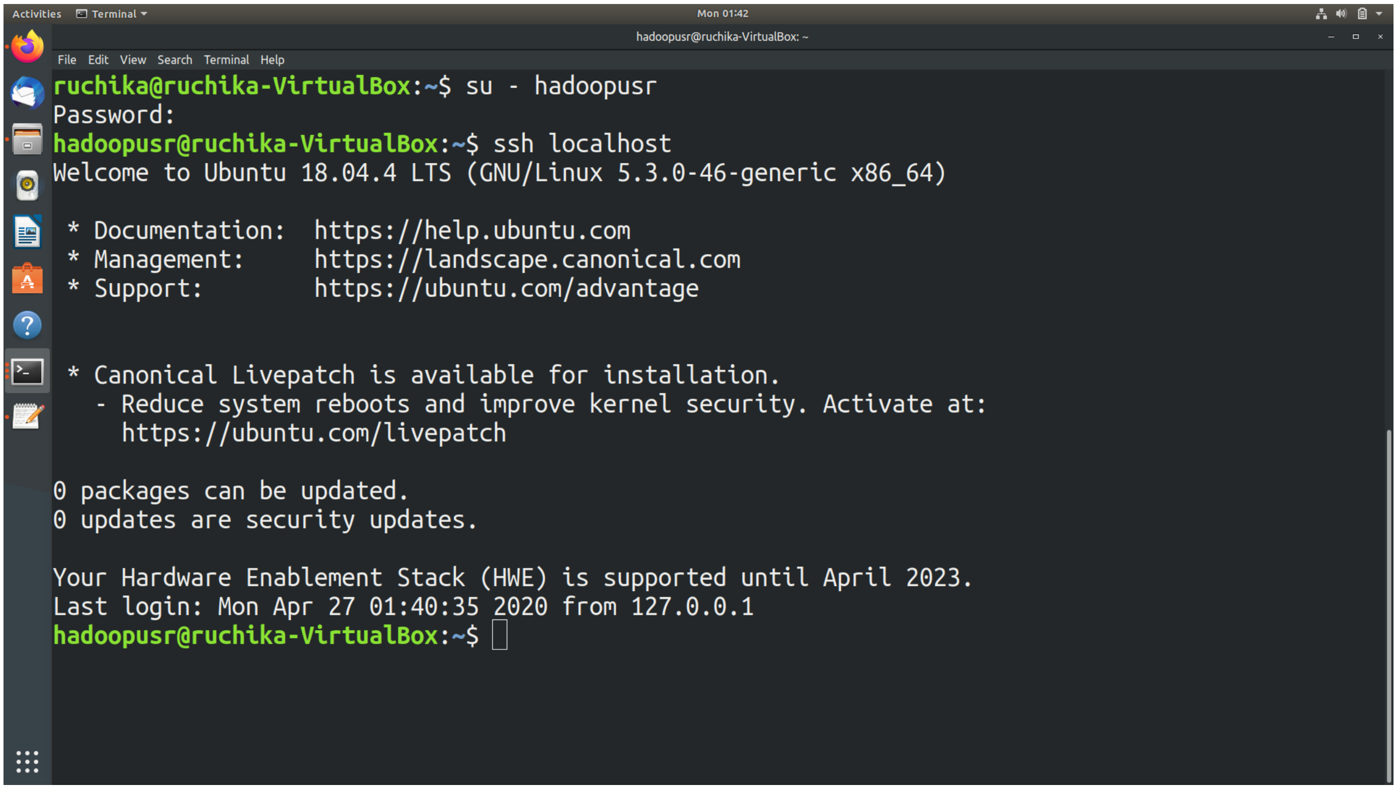
- Run the command below to produce *data.txt*

  ```
  python3 /preprocessing/cv_green.py
  ```

Steps to run Map-Reduce Job:

- Install Hadoop on Ubuntu using the guide here
- Once the installation is complete, you will have two users on your system, ***username*** and hadoopusr.
- Login into ***username*** using your password.
- Open Terminal in the user: ***username***
- Use the following command to login to the hadoopusr and then ssh into localhost:

```
su - hadoopusr
ssh localhost
```



Activities
Terminal
Mon 01:42
hadoopusr@ruchika-VirtualBox: ~
File Edit View Search Terminal Help

```
ruchika@ruchika-VirtualBox:~$ su - hadoopusr
Password:
hadoopusr@ruchika-VirtualBox:~$ ssh localhost
Welcome to Ubuntu 18.04.4 LTS (GNU/Linux 5.3.0-46-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage


 * Canonical Livepatch is available for installation.
   - Reduce system reboots and improve kernel security. Activate at:
     https://ubuntu.com/livepatch

0 packages can be updated.
0 updates are security updates.

Your Hardware Enablement Stack (HWE) is supported until April 2023.
Last login: Mon Apr 27 01:40:35 2020 from 127.0.0.1
hadoopusr@ruchika-VirtualBox:~$ 
```

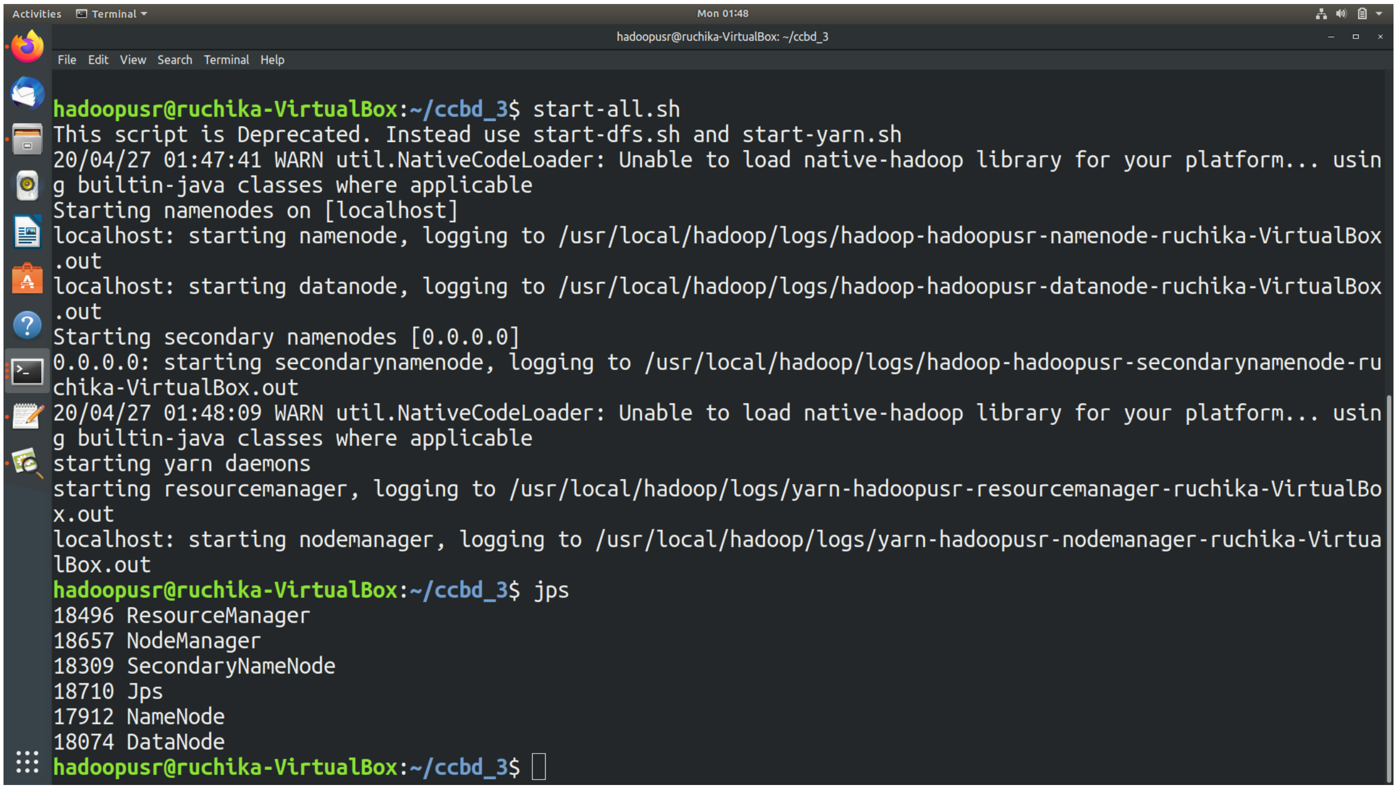
- Enter password on prompt
- Run the following command to cd into the folder:

  `cd /home/username/ccbd_3`

- To start all nodes:

```
start-all.sh
```

- To check whether all nodes are running, use the `jps` command. The output should look similar to the image below.


Activities
Terminal
Mon 01:48
hadoopusr@ruchika-VirtualBox: ~/ccbd_3
File Edit View Search Terminal Help

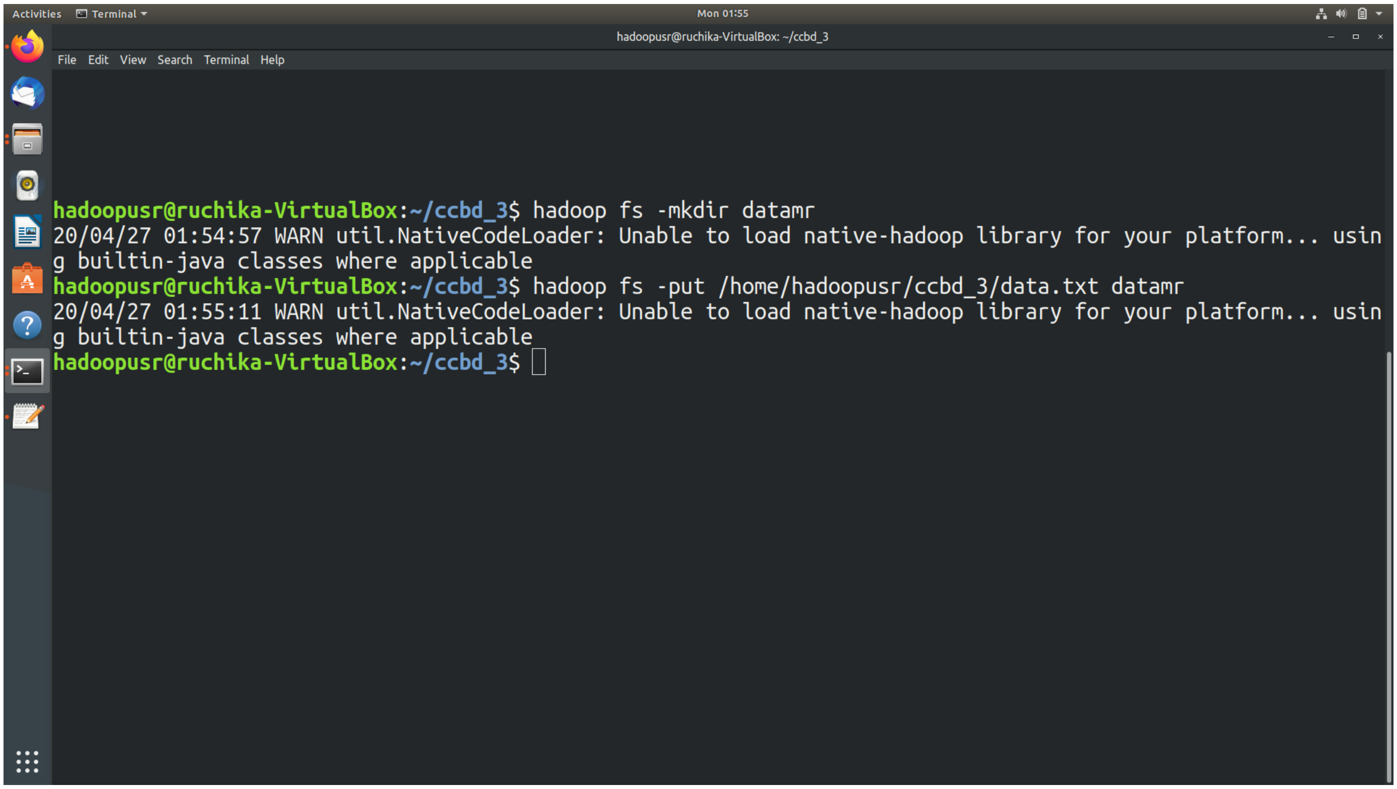
```
hadoopusr@ruchika-VirtualBox:~/ccbd_3$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
20/04/27 01:47:41 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... usin
g builtin-java classes where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hadoopusr-namenode-ruchika-VirtualBox
.out
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hadoopusr-datanode-ruchika-VirtualBox
.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-hadoopusr-secondarynamenode-ru
chika-VirtualBox.out
20/04/27 01:48:09 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... usin
g builtin-java classes where applicable
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-hadoopusr-resourcemanager-ruchika-VirtualBo
x.out
localhost: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-hadoopusr-nodemanager-ruchika-Virtua
lBox.out
hadoopusr@ruchika-VirtualBox:~/ccbd_3$ jps
18496 ResourceManager
18657 NodeManager
18309 SecondaryNameNode
18710 Jps
17912 NameNode
18074 DataNode
hadoopusr@ruchika-VirtualBox:~/ccbd_3$ 
```

- To make a folder in file system on hadoop, use the command below:

```
hadoop fs -mkdir /datamr
```

- To copy data.txt from local system to hadoop file system , use the command below:

```
hadoop fs -put /home/username/ccbd_3/data.txt /datamr
```


Activities
Terminal
Mon 01:55
hadoopusr@ruchika-VirtualBox: ~/ccbd_3
File Edit View Search Terminal Help

```
hadoopusr@ruchika-VirtualBox:~/ccbd_3$ hadoop fs -mkdir datamr
20/04/27 01:54:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... usin
g builtin-java classes where applicable
hadoopusr@ruchika-VirtualBox:~/ccbd_3$ hadoop fs -put /home/hadoopusr/ccbd_3/data.txt datamr
20/04/27 01:55:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... usin
g builtin-java classes where applicable
hadoopusr@ruchika-VirtualBox:~/ccbd_3$ 
```

- To confirm whether its done or not, use the following command

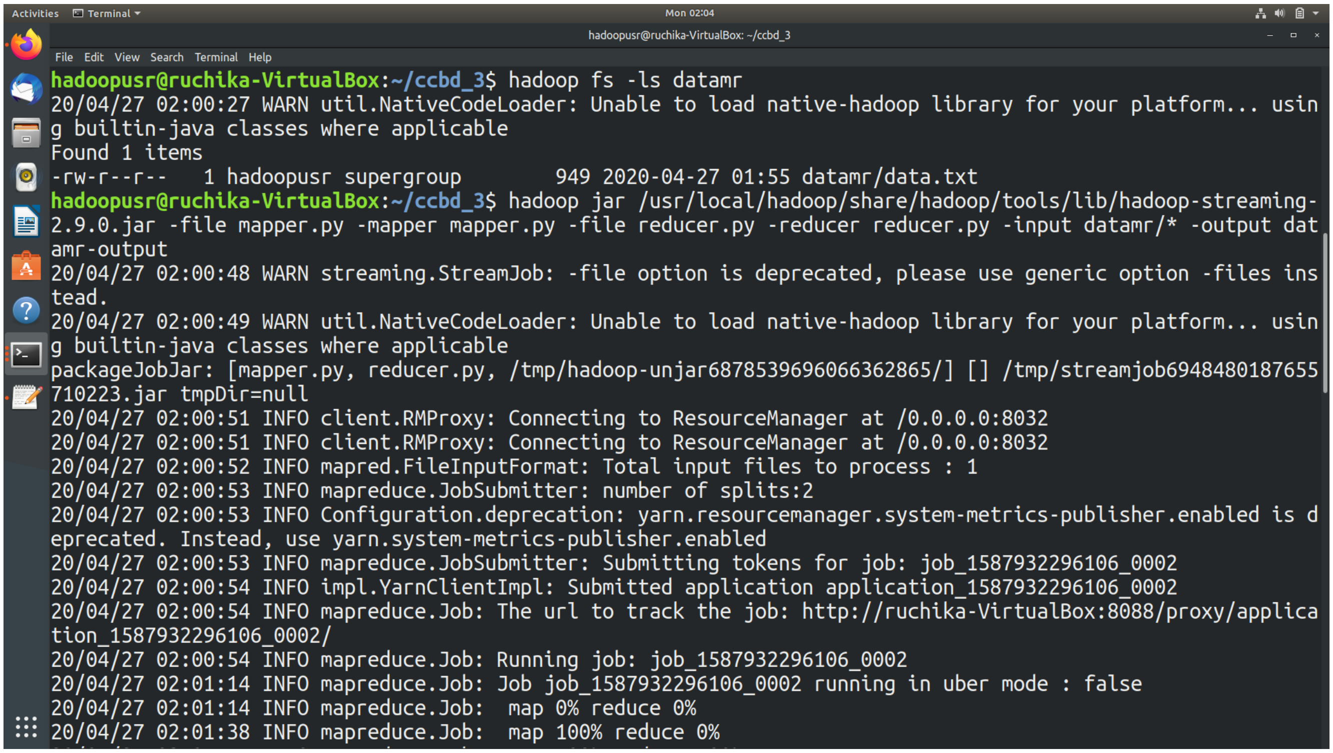
```
hadoop fs -ls /datamr
```

- The output of this command should look something like this.
- To run the task:

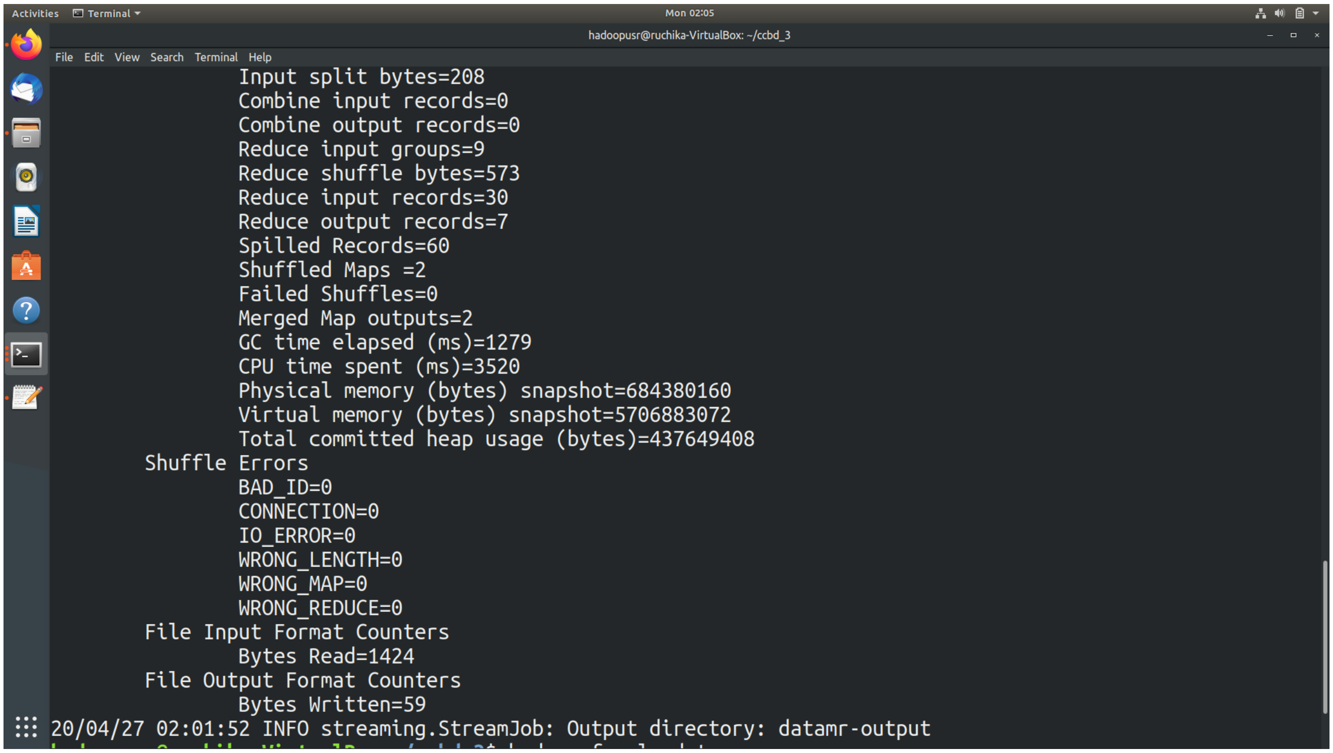
```
hadoop jar /usr/local/hadoop/share/hadoop/tools/lib/hadoop-streaming-2.9.0.jar -file mapper.py -map
```

  - You can use the following flag to define the number of reduce tasks explicitly:

    ```
    -D mapred.reduce.tasks=<number-of-tasks>
    ```


Activities
Terminal
Mon 02:04
hadoopusr@ruchika-VirtualBox: ~/ccbd_3
File Edit View Search Terminal Help

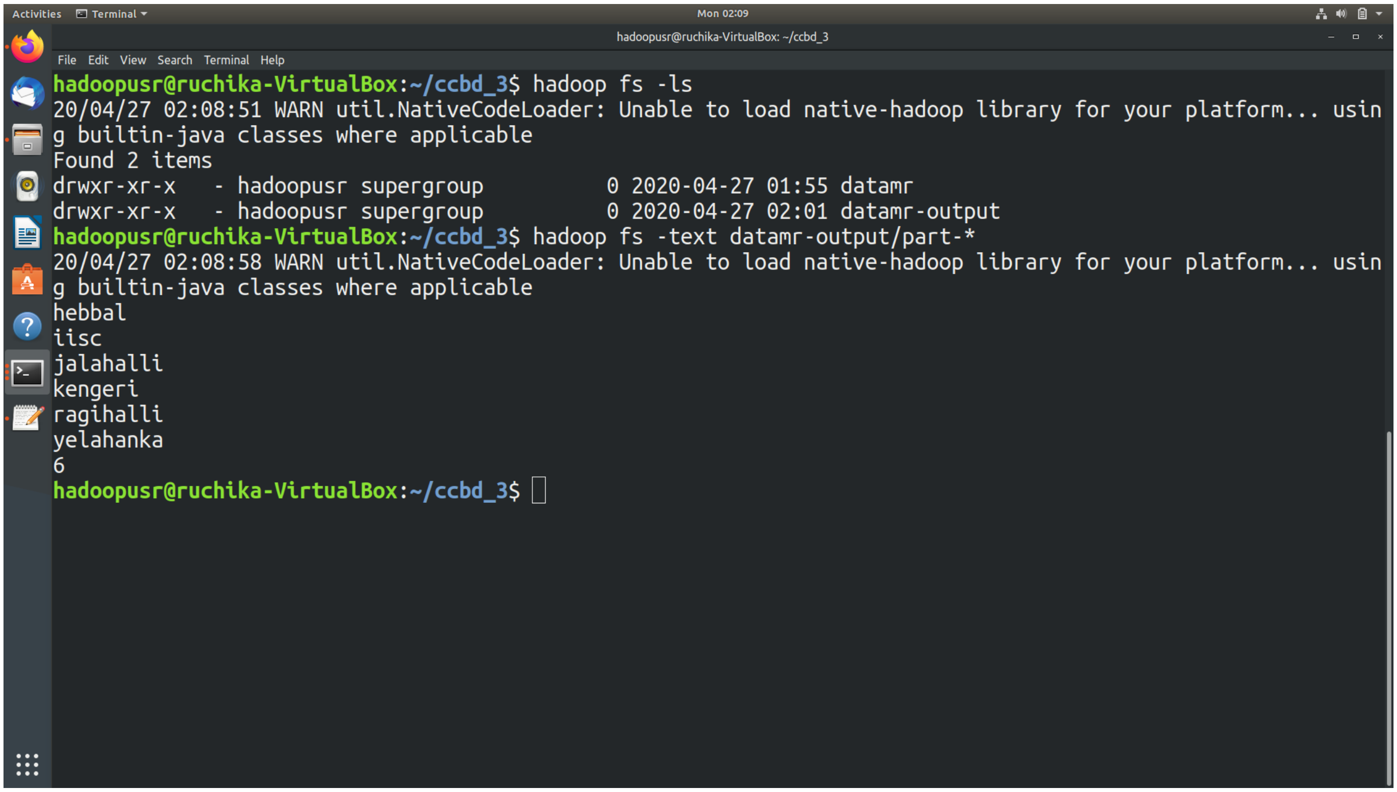
```
hadoopusr@ruchika-VirtualBox:~/ccbd_3$ hadoop fs -ls datamr
20/04/27 02:00:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... usin
g builtin-java classes where applicable
Found 1 items
-rw-r--r--   1 hadoopusr supergroup        949 2020-04-27 01:55 datamr/data.txt
hadoopusr@ruchika-VirtualBox:~/ccbd_3$ hadoop jar /usr/local/hadoop/share/hadoop/tools/lib/hadoop-streaming-
2.9.0.jar -file mapper.py -mapper mapper.py -file reducer.py -reducer reducer.py -input datamr/* -output dat
amr-output
20/04/27 02:00:48 WARN streaming.StreamJob: -file option is deprecated, please use generic option -files ins
tead.
20/04/27 02:00:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... usin
g builtin-java classes where applicable
packageJobJar: [mapper.py, reducer.py, /tmp/hadoop-unjar6878539696066362865/] [] /tmp/streamjob6948480187655
710223.jar tmpDir=null
20/04/27 02:00:51 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
20/04/27 02:00:51 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
20/04/27 02:00:52 INFO mapred.FileInputFormat: Total input files to process : 1
20/04/27 02:00:53 INFO mapreduce.JobSubmitter: number of splits:2
20/04/27 02:00:53 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is d
eprecated. Instead, use yarn.system-metrics-publisher.enabled
20/04/27 02:00:53 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1587932296106_0002
20/04/27 02:00:54 INFO impl.YarnClientImpl: Submitted application application_1587932296106_0002
20/04/27 02:00:54 INFO mapreduce.Job: The url to track the job: http://ruchika-VirtualBox:8088/proxy/applica
tion_1587932296106_0002/
20/04/27 02:00:54 INFO mapreduce.Job: Running job: job_1587932296106_0002
20/04/27 02:01:14 INFO mapreduce.Job: Job job_1587932296106_0002 running in uber mode : false
20/04/27 02:01:14 INFO mapreduce.Job:  map 0% reduce 0%
20/04/27 02:01:38 INFO mapreduce.Job:  map 100% reduce 0%
```


Activities
Terminal
Mon 02:05
hadoopusr@ruchika-VirtualBox: ~/ccbd_3
File Edit View Search Terminal Help

```
                Input split bytes=208
                Combine input records=0
                Combine output records=0
                Reduce input groups=9
                Reduce shuffle bytes=573
                Reduce input records=30
                Reduce output records=7
                Spilled Records=60
                Shuffled Maps =2
                Failed Shuffles=0
                Merged Map outputs=2
                GC time elapsed (ms)=1279
                CPU time spent (ms)=3520
                Physical memory (bytes) snapshot=684380160
                Virtual memory (bytes) snapshot=5706883072
                Total committed heap usage (bytes)=437649408
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=1424
        File Output Format Counters
                Bytes Written=59
20/04/27 02:01:52 INFO streaming.StreamJob: Output directory: datamr-output
```

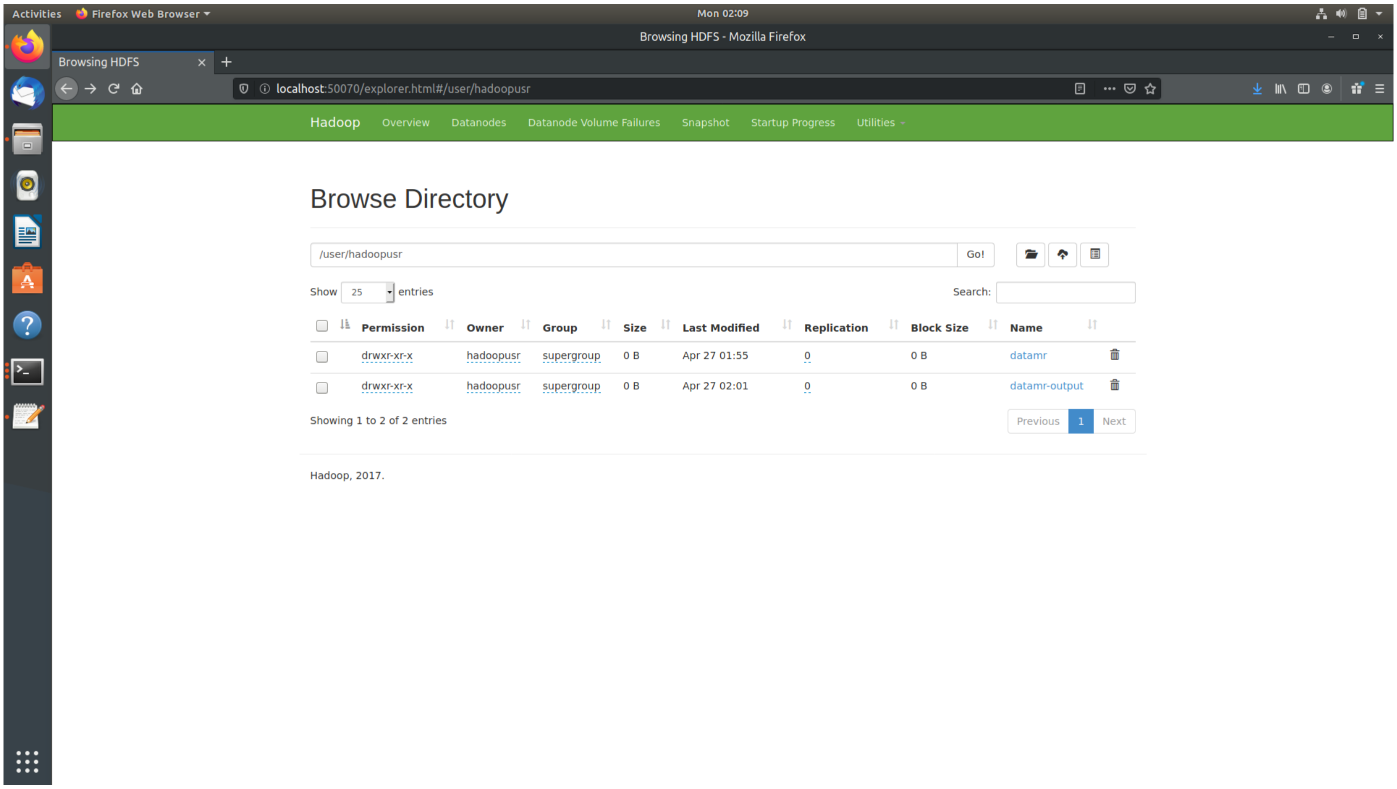
- Use the following command to access the output:

```
hadoop fs -text /datamr-output/part-*
```


Activities
Terminal
Mon 02:09
hadoopusr@ruchika-VirtualBox: ~/ccbd_3
File Edit View Search Terminal Help

```
hadoopusr@ruchika-VirtualBox:~/ccbd_3$ hadoop fs -ls
20/04/27 02:08:51 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... usin
g builtin-java classes where applicable
Found 2 items
drwxr-xr-x   - hadoopusr supergroup          0 2020-04-27 01:55 datamr
drwxr-xr-x   - hadoopusr supergroup          0 2020-04-27 02:01 datamr-output
hadoopusr@ruchika-VirtualBox:~/ccbd_3$ hadoop fs -text datamr-output/part-*
20/04/27 02:08:58 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... usin
g builtin-java classes where applicable
hebbal
iisc
jalahalli
kengeri
ragihalli
yelahanka
6
hadoopusr@ruchika-VirtualBox:~/ccbd_3$ 
```

- Visit `localhost:50070/explorer.html/user/hadoopusr` to view the job result, the result should look similar to the image below.



Browse Directory

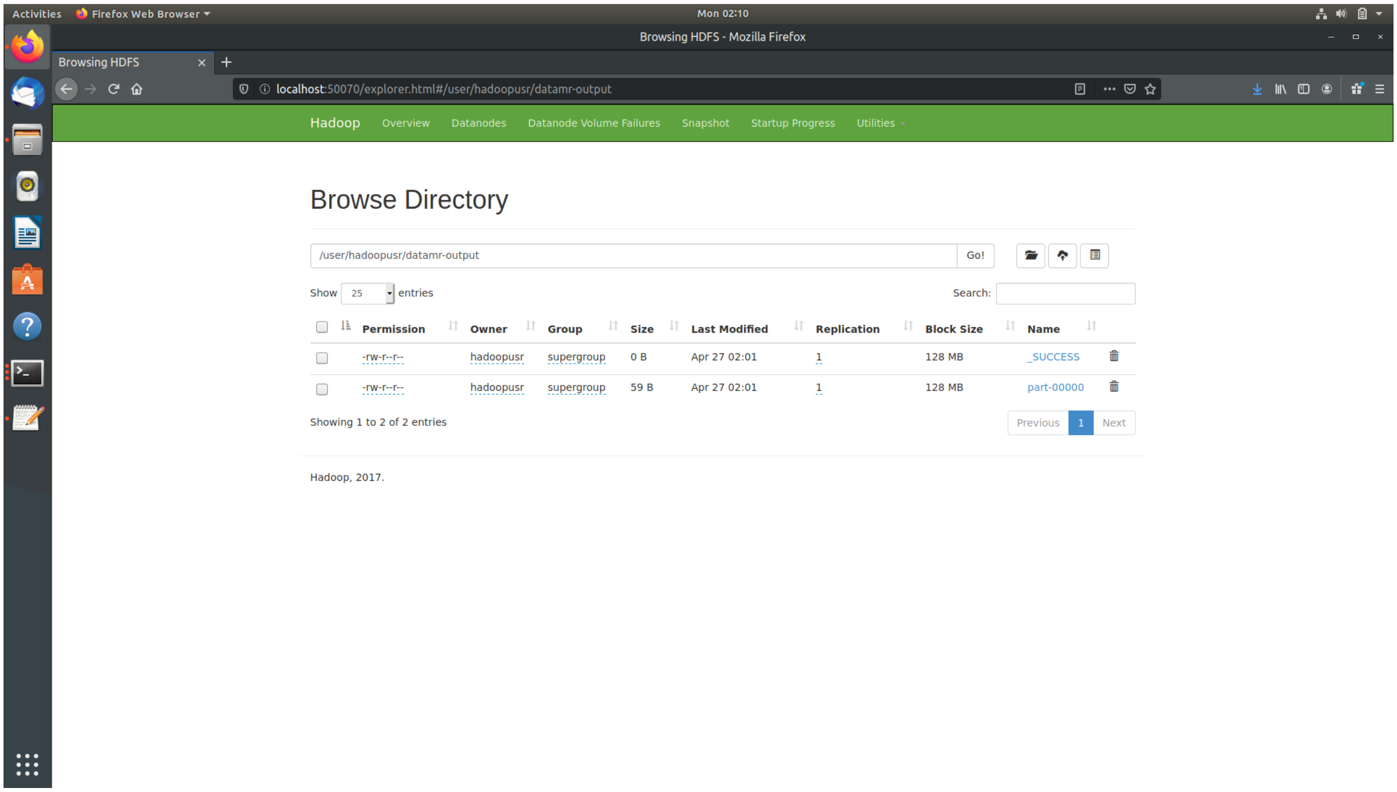
/user/hadoopusr

□	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
□	drwxr-xr-x	hadoopusr	supergroup	0 B	Apr 27 01:55	0	0 B	datamr
□	drwxr-xr-x	hadoopusr	supergroup	0 B	Apr 27 02:01	0	0 B	datamr-output

Showing 1 to 2 of 2 entries

Hadoop, 2017.

- Click on `data-output` and then click on `part-00000`



Browse Directory

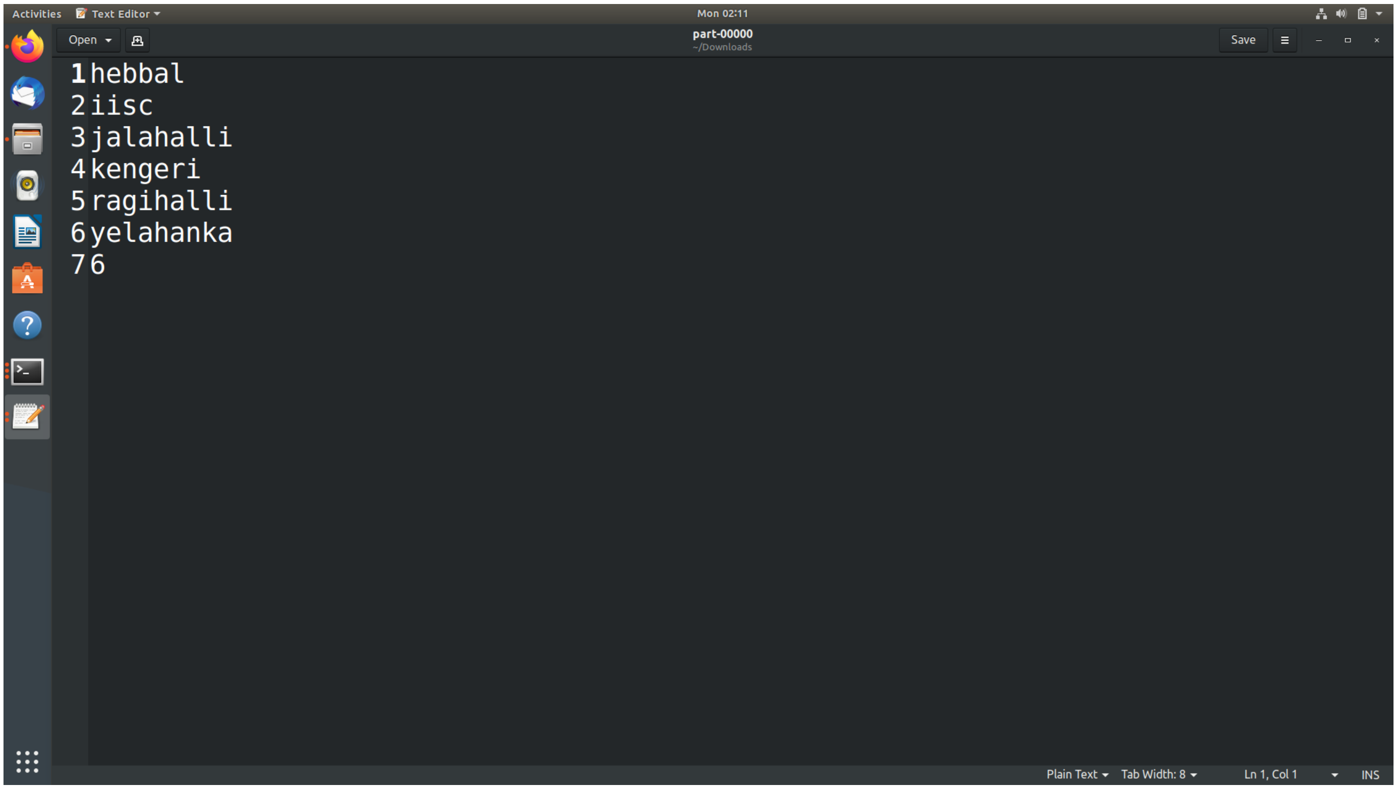
/user/hadoopusr/datamr-output

□	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
□	-rw-r--r--	hadoopusr	supergroup	0 B	Apr 27 02:01	1	128 MB	_SUCCESS
□	-rw-r--r--	hadoopusr	supergroup	59 B	Apr 27 02:01	1	128 MB	part-00000

Showing 1 to 2 of 2 entries

Hadoop, 2017.

- Press Download to fetch a copy of the result of the map-reduce job
- Open the file to view the result.



```
hebbal
iisc
jalahalli
kengeri
ragihalli
yelahanka
6
```

Note: File structure and port numbers may vary depending on the installation method.

Useful Commads:

- Hadoop:
  - Start all nodes: `start-all/sh`
  - Stop all nodes: `stop-all.sh`
  - Format Namenode: `hdfs namenode -format`
  - Java Virtual Machine Process Status Tool: `jps`
- HDFS:
  - List all Directories and Files:

    ```
    hadoop dfs -ls /path/to/directory
    ```
  - To make a folder in HDFS:

    ```
    hadoop fs -mkdir /path/to/folder
    ```
  - To copy a folder from local system to HDFS:

    ```
    hadoop fs put /path/to/local/directory /path/to/hdfs/directory
    ```
  - To delete a directory in HDFS:

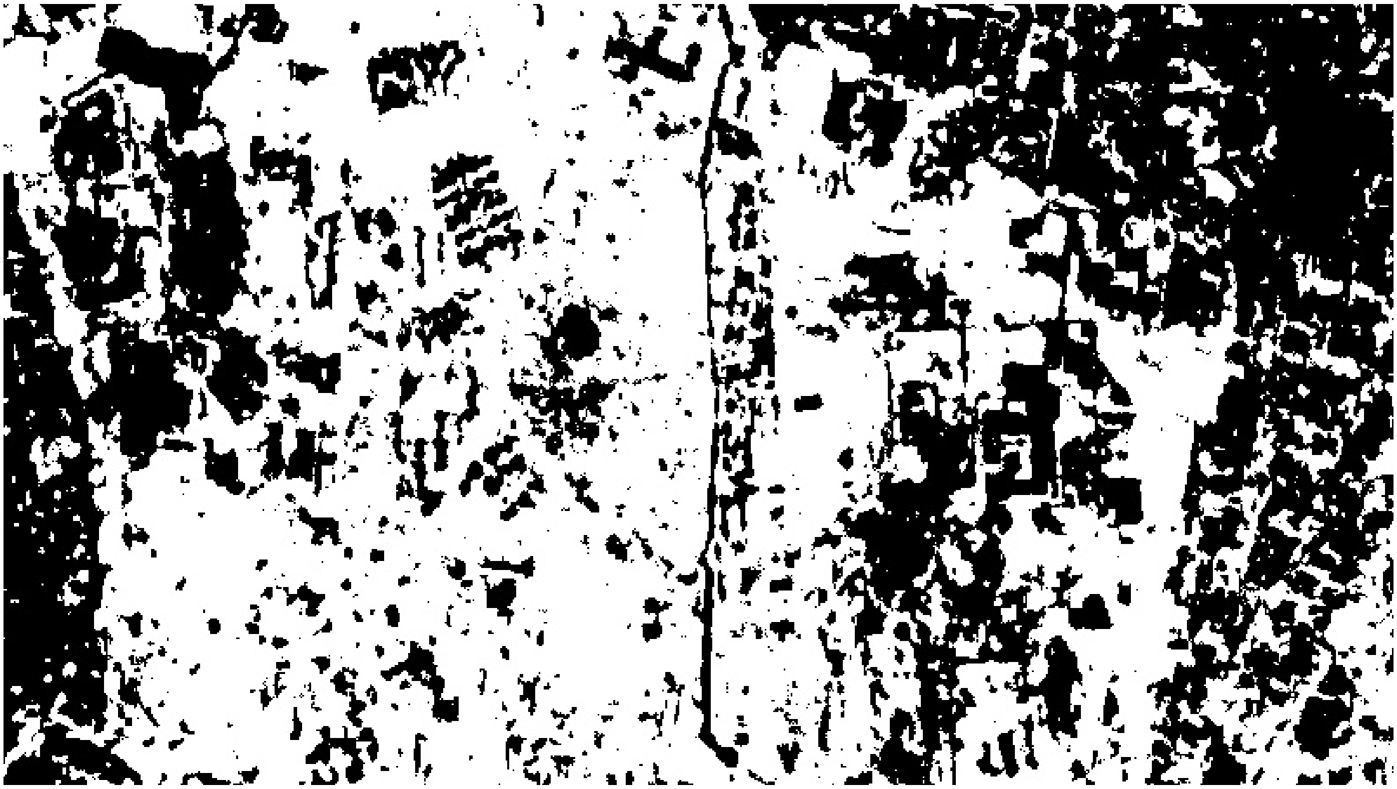
    ```
    hadoop fs -rm -r /path/to/directory
    ```
  - To view result of a Job:

    ```
    hadoop fs -text /path/to/output/part-*
    ```

Explaination:

Once the cv_green.py runs, these steps take place iteratively for each image in the /preprocessing/images folder:

- The image is resized to a 16:9 ratio.
- Every single pixel is checked whether it has more green content in it compared to the other two, if yes - it is replaced with a white pixel else with a black pixel.
- At the end of the previous step, we get a black and white image where every green pixel is mapped to white and rest to black.
- The issue of using this however is that this approach avoids the dark spots in between the trees in the images, so to avoid this, we use the dilate method in open-cv which helps in covering these dark spots in between the trees.
- An example output image can be seen below:



- In this image, the black pixel clearly represent box-shaped house like structures.
- The number of white pixels divided by the total number of pixels in the image gives us the percentage greenery in the image.

Once this is done, we need to consider the fact that each area has multiple images. So we calculate the average greenery among all these images for the given area and consider this to be the final greenery percentage of that area. However, this can be used only when the scale of all the images are the same and this was taken care of while collecting these Satellite images.