Problem Statement 1

1. Create a folder in HDFS by name "dir01" and move input1.txt , input2.txt and input3.txt into /dir01.

```
[march8lab23@ip-10-1-1-204 workdir]$ pwd
/home/march8lab23/big-data-developer-course/workdir
[march8lab23@ip-10-1-1-204 workdir]$ hdfs dfs -mkdir /user/$USER/soumya/dir01
[march8lab23@ip-10-1-1-204 workdir]$ hdfs dfs -ls /user/$USER/soumya
Found 6 items
drwxr-xr-x   - march8lab23 march8lab23          0 2023-07-24 11:17 /user/march8lab23/soumya/dir01
-rw-r--r--   5 march8lab23 march8lab23         37 2023-07-24 10:33 /user/march8lab23/soumya/file1.txt
-rw-r--r--   5 march8lab23 march8lab23         37 2023-07-24 10:35 /user/march8lab23/soumya/file2.txt
-rw-r--r--   3 march8lab23 march8lab23       9508 2023-07-24 10:54 /user/march8lab23/soumya/input1.txt
-rw-r--r--   3 march8lab23 march8lab23      19718 2023-07-24 10:54 /user/march8lab23/soumya/input2.txt
-rw-r--r--   3 march8lab23 march8lab23      30692 2023-07-24 10:54 /user/march8lab23/soumya/input3.txt
```

2. List only the file names present in "/dir01"

```
[march8lab23@ip-10-1-1-204 workdir]$ cd
[march8lab23@ip-10-1-1-204 ~]$ ls -lrt
total 256
-rwxrwxrwx 1 march8lab23 march8lab23 142551 Apr 18 09:11 hadoop-examples.jar
-rw-r--r-- 1 march8lab23 march8lab23   1727 Apr 24 09:39 Untitled1.ipynb
-rw-r--r-- 1 march8lab23 march8lab23   7946 Apr 24 09:44 Untitled.ipynb
-rw-r--r-- 1 march8lab23 march8lab23   4715 Apr 24 10:54 Untitled2.ipynb
drwxrwxr-x 2 march8lab23 march8lab23   4096 May 16 12:13 pushpalatha
drwxr-xr-- 2 march8lab23 march8lab23   4096 Jun 12 12:09 venkatesh_chavan
drwxr-xr-- 2 march8lab23 march8lab23   4096 Jun 12 12:09 sai_koteswarachowdary
drwxr-xr-- 2 march8lab23 march8lab23   4096 Jun 12 12:10 arulprakash
drwxr-xr-- 2 march8lab23 march8lab23   4096 Jun 12 12:11 swathi_shivamadhu
drwxr-xr-- 2 march8lab23 march8lab23   4096 Jun 12 12:36 sivaiah_guggilla
-rw-rw-r-- 1 march8lab23 march8lab23      0 Jun 19 12:01 hdfs
-rw-rw-r-- 1 march8lab23 march8lab23    348 Jun 19 12:04 CREATE
-rw-rw-r-- 1 march8lab23 march8lab23    357 Jun 19 12:11 USE
-rw-rw-r-- 1 march8lab23 march8lab23    162 Jun 22 07:37 employee1.dat
drwxrwxr-x 3 march8lab23 march8lab23   4096 Jun 26 11:25 big-data-developer-course_bkp
drwxrwxr-x 7 march8lab23 march8lab23   4096 Jun 26 11:26 big-data-developer-course
-rw-r--r-- 1 march8lab23 march8lab23    555 Jul  6 09:18 Untitled3.ipynb
-rw-r--r-- 1 march8lab23 march8lab23    555 Jul  6 09:18 Untitled3-Copy1.ipynb
-rw-r--r-- 1 march8lab23 march8lab23    555 Jul  6 09:18 Sandhya_Shankar.ipynb
-rw-rw-r-- 1 march8lab23 march8lab23      0 Jul  7 11:34 sqoop
drwxrwxr-x 3 march8lab23 march8lab23   4096 Jul 10 13:53 lokesh
-rw-r--r-- 1 march8lab23 march8lab23  12124 Jul 10 15:20 Untitled4.ipynb
-rw-r--r-- 1 march8lab23 march8lab23    284 Jul 15 13:51 input1.json
drwxrwxr-x 5 march8lab23 march8lab23   4096 Jul 17 07:40 sandhya
drwxrwxr-x 5 march8lab23 march8lab23   4096 Jul 18 13:18 mahesh
drwxrwxr-x 3 march8lab23 march8lab23   4096 Jul 19 07:50 vasagirir
drwxr-xr-x 6 march8lab23 march8lab23   4096 Jul 24 09:47 Mahesh
drwxrwxr-x 2 march8lab23 march8lab23   4096 Jul 24 10:51 soumya
[march8lab23@ip-10-1-1-204 ~]$ cd soumya
[march8lab23@ip-10-1-1-204 soumya]$ ls -lrt
total 64
-rw-rw-r-- 1 march8lab23 march8lab23  9508 Jul 24 10:50 input1.txt
-rw-rw-r-- 1 march8lab23 march8lab23 19718 Jul 24 10:51 input2.txt
-rw-rw-r-- 1 march8lab23 march8lab23 30692 Jul 24 10:52 input3.txt
[march8lab23@ip-10-1-1-204 soumya]$ hdfs dfs -put input*.txt /user/$USER/soumya/dir01
[march8lab23@ip-10-1-1-204 soumya]$ hdfs dfs -ls /user/$USER/soumya/dir01
Found 3 items
-rw-r--r--   3 march8lab23 march8lab23       9508 2023-07-24 11:21 /user/march8lab23/soumya/dir01/input1.t
xt
-rw-r--r--   3 march8lab23 march8lab23      19718 2023-07-24 11:21 /user/march8lab23/soumya/dir01/input2.t
xt
-rw-r--r--   3 march8lab23 march8lab23      30692 2023-07-24 11:21 /user/march8lab23/soumya/dir01/input3.t
xt
```

3. Change the replication factor for the content present in directory "/dir01" to 5 and display the replication factor for the files present in "/dir01".

```
[march8lab23@ip-10-1-1-204 soumya]$ hdfs dfs -setrep 5 /user/$USER/soumya/dir01
Replication 5 set: /user/march8lab23/soumya/dir01/input1.txt
Replication 5 set: /user/march8lab23/soumya/dir01/input2.txt
Replication 5 set: /user/march8lab23/soumya/dir01/input3.txt
```

4. Create a folder in HDFS by name "scenario01" and create directory "level01" inside "scenario01" directory and create another directory "level02" inside directory "level01". Once the required directories are created copy input1.txt to scenario01, input2.txt to level01 and input3.txt to level02 and finally recursively print only the file names present in directory scenario01 /user/naveen/scenario01/input1.txt /user/naveen/scenario01/level01 /user/naveen/scenario01/level01/input1.txt /user/naveen/scenario01/level01/level02 /user/naveen/scenario01/level02 /user/naveen/scenario01/level03

```
[march8lab23@ip-10-1-1-204 soumya]$ hdfs dfs -mkdir /user/$USER/soumya/scenario01
[march8lab23@ip-10-1-1-204 soumya]$ hdfs dfs -mkdir /user/$USER/soumya/scenario01/level01
[march8lab23@ip-10-1-1-204 soumya]$ hdfs dfs -mkdir /user/$USER/soumya/scenario01/level01/level02
[march8lab23@ip-10-1-1-204 soumya]$ hdfs dfs -ls /user/$USER/soumya
Found 7 items
drwxr-xr-x   - march8lab23 march8lab23          0 2023-07-24 11:21 /user/march8lab23/soumya/dir01
-rw-r--r--   5 march8lab23 march8lab23         37 2023-07-24 10:33 /user/march8lab23/soumya/file1.txt
-rw-r--r--   5 march8lab23 march8lab23         37 2023-07-24 10:35 /user/march8lab23/soumya/file2.txt
-rw-r--r--   3 march8lab23 march8lab23       9508 2023-07-24 10:54 /user/march8lab23/soumya/input1.txt
-rw-r--r--   3 march8lab23 march8lab23      19718 2023-07-24 10:54 /user/march8lab23/soumya/input2.txt
-rw-r--r--   3 march8lab23 march8lab23      30692 2023-07-24 10:54 /user/march8lab23/soumya/input3.txt
drwxr-xr-x   - march8lab23 march8lab23          0 2023-07-24 11:43 /user/march8lab23/soumya/scenario01
[march8lab23@ip-10-1-1-204 soumya]$ hdfs dfs -cp input1.txt /user/$USER/soumya/scenario01
cp: `input1.txt': No such file or directory
[march8lab23@ip-10-1-1-204 soumya]$ hdfs dfs -put input1.txt /user/$USER/soumya/scenario01
[march8lab23@ip-10-1-1-204 soumya]$ hdfs dfs -put input2.txt /user/$USER/soumya/scenario01/level01
[march8lab23@ip-10-1-1-204 soumya]$ hdfs dfs -put input3.txt /user/$USER/soumya/scenario01/level01/level02
[march8lab23@ip-10-1-1-204 soumya]$ hdfs dfs -ls /user/$USER/scenario02
ls: `/user/march8lab23/scenario02': No such file or directory
[march8lab23@ip-10-1-1-204 soumya]$ hdfs dfs -ls /user/$USER/scenario01
ls: `/user/march8lab23/scenario01': No such file or directory
[march8lab23@ip-10-1-1-204 soumya]$ hdfs dfs -ls /user/$USER/soumya/scenario01
Found 2 items
-rw-r--r--   3 march8lab23 march8lab23       9508 2023-07-24 11:47 /user/march8lab23/soumya/scenario01/inp
ut1.txt
drwxr-xr-x   - march8lab23 march8lab23          0 2023-07-24 11:47 /user/march8lab23/soumya/scenario01/lev
el01
```

Problem Statement 2

Working with YARN Commands Commands

1. Run MapReduce Program and capture the application Id of the job.

```
[march8lab23@ip-10-1-1-204 ~]$ cd /opt/cloudera/parcels/CDH/jars
[march8lab23@ip-10-1-1-204 jars]$ ls hadoop-mapreduce-examples-3.0.0-cdh6.2.1.jar
hadoop-mapreduce-examples-3.0.0-cdh6.2.1.jar
[march8lab23@ip-10-1-1-204 jars]$ hdfs dfs -cat /user/$USER/soumya/*.txt
-----File 1.txt-----
Line 01
Line 02
-----File 2.txt-----
Line 01
Line 02
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vivamus condimentum sagittis lacus, laoreet luctu
s ligula laoreet ut. Vestibulum ullamcorper accumsan velit vel vehicula. Proin tempor lacus arcu. Nunc at
elit condimentum, semper nisi et, condimentum mi. In venenatis blandit nibh at sollicitudin. Vestibulum da
pibus mauris at orci maximus pellentesque. Nullam id elementum ipsum. Suspendisse cursus lobortis viverra.
 Proin et erat at mauris tincidunt porttitor vitae ac dui.


Donec vulputate lorem tortor, nec fermentum nibh bibendum vel. Lorem ipsum dolor sit amet, consectetur adi
piscing elit. Praesent dictum luctus massa, non euismod lacus. Pellentesque condimentum dolor est, ut dapi
bus lectus luctus ac. Ut sagittis commodo arcu. Integer nisi nulla, facilisis sit amet nulla quis, eleifen
d suscipit purus. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenae
```

```
[march8lab23@ip-10-1-1-204 jars]$ yarn jar hadoop-mapreduce-examples-3.0.0-cdh6.2.1.jar wordcount /user/$U
SER/soumya/   /user/$USER/soumya_output/
WARNING: YARN_OPTS has been replaced by HADOOP_OPTS. Using value of YARN_OPTS.
23/07/24 12:40:21 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.i
nternal/10.1.1.204:8032
23/07/24 12:40:21 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /user/march8lab23
/.staging/job_1685754149182_7428
23/07/24 12:40:21 INFO input.FileInputFormat: Total input files to process : 7
23/07/24 12:40:22 INFO mapreduce.JobSubmitter: number of splits:7
23/07/24 12:40:22 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is
 deprecated. Instead, use yarn.system-metrics-publisher.enabled
23/07/24 12:40:22 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1685754149182_7428
23/07/24 12:40:22 INFO mapreduce.JobSubmitter: Executing with tokens: []
23/07/24 12:40:22 INFO conf.Configuration: resource-types.xml not found
23/07/24 12:40:22 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
23/07/24 12:40:22 INFO impl.YarnClientImpl: Submitted application application_1685754149182_7428
```

2. Re-run the MapReduce program and kill the application using the yarn command.

```
[march8lab23@ip-10-1-1-204 jars]$ yarn jar hadoop-mapreduce-examples-3.0.0-cdh6.2.1.jar wordcount /user/$U
SER/soumya/ /user/$USER/soumyaba_output/
WARNING: YARN_OPTS has been replaced by HADOOP_OPTS. Using value of YARN_OPTS.
23/07/24 12:55:03 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.i
nternal/10.1.1.204:8032
23/07/24 12:55:04 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /user/march8lab23
/.staging/job_1685754149182_7430
23/07/24 12:55:04 INFO input.FileInputFormat: Total input files to process : 7
23/07/24 12:55:04 INFO mapreduce.JobSubmitter: number of splits:7
23/07/24 12:55:05 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is
 deprecated. Instead, use yarn.system-metrics-publisher.enabled
23/07/24 12:55:05 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1685754149182_7430
23/07/24 12:55:05 INFO mapreduce.JobSubmitter: Executing with tokens: []
23/07/24 12:55:05 INFO conf.Configuration: resource-types.xml not found
23/07/24 12:55:05 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
23/07/24 12:55:05 INFO impl.YarnClientImpl: Submitted application application_1685754149182_7430
23/07/24 12:55:05 INFO mapreduce.Job: The url to track the job: http://ip-10-1-1-204.ap-south-1.compute.in
ternal:6066/proxy/application_1685754149182_7430/
23/07/24 12:55:05 INFO mapreduce.Job: Running job: job_1685754149182_7430
^C[march8lab23@ip-10-1-1-204 jars]$ mapred job -kill 1685754149182
23/07/24 12:56:07 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.i
nternal/10.1.1.204:8032
```

3. List all the applications which are RUNNING state

```
[march8lab23@ip-10-1-1-204 jars]$ yarn application -list -appStates RUNNING
WARNING: YARN_OPTS has been replaced by HADOOP_OPTS. Using value of YARN_OPTS.
23/07/24 12:58:43 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.i
nternal/10.1.1.204:8032
Total number of applications (application-types: [], states: [RUNNING] and tags: []):3
                Application-Id      Application-Name      Application-Type          User          Queue
            State          Final-State          Progress                      Tracking-URL
application_1685754149182_7150          HdfsTest_py              SPARK    ssasidaran73edu root.defau
lt            RUNNING              UNDEFINED              10% http://ip-10-1-1-204.ap-south-1.co
mpute.internal:18088/history/application_1685754149182_7150/1
application_1685754149182_7429  bike_rental_prediction              SPARK    veronica1167edu root.defau
lt            RUNNING              UNDEFINED              10% http://ip-10-1-1-204.ap-south-1.co
mpute.internal:43024
application_1685754149182_7427          PySparkShell              SPARK    nithins4tc13edu root.defau
lt            RUNNING              UNDEFINED              10% http://ip-10-1-1-204.ap-south-1.co
mpute.internal:18088/history/application_1685754149182_7427/1
```

4. View the logs of any of the jobs which are already completed.

yarn application -list -appStates FINISHED