# **HADOOP ASSIGNMENT**

**Problem Statement 1:**

1. hdfs dfs -mkdir /dir01 #create directory dir01

hdfs dfs -put input1.txt input2.txt input3.txt /dir01 #Put files in directory

1. hdfs dfs -ls /dir01 #Files in dir01
2. hdfs dfs -setrep 5 /user/$USER/dir01 #Set replication factor to 5 for dir01

hdfs dfs -stat %r /dir01 # Get replication factor for dir01

1. hdfs dfs -mkdir /scenario01 # #create directory scenario01

hdfs dfs -mkdir /scenario01/level01 # Create level01 directory inside scenario01

hdfs dfs -mkdir /scenario01/level01/level02 # Create level02 directory inside level01

hdfs dfs -put input1.txt /scenario01 # Copy input1.txt to scenario01

hdfs dfs -put input2.txt /scenario01/level01 # Copy input2.txt to level01

hdfs dfs -put input3.txt /scenario01/level01/level02 # Copy "input3.txt" to "level02"

hdfs dfs -ls -R /scenario01 # Recursively print only the file names present in scenario01

**YARN Commands**

**Problem Statemen 2:**

1. # Capture the application id for the job

yarn jar hadoop-mapreduce-jar.jar APPLICATION\_ID /user/dir01/input / /user/dir01/output/

1. # Re-run map-reduce program

yarn jar hadoop-mapreduce-jar.jar APPLICATION\_ID /user/dir01/input / /user/dir01/output/

# Kill the application

yarn application -kill <Application\_ID>

1. yarn application -list -appStates RUNNING # list applications in running state
2. yarn logs -applicationId < Application\_ID > # view logs of jobs completed