# Program JAR Files

## All the required jar files are available in the Lab @ the following path -

## /data/trendytech/mapreduce_jars

**Note : These JAR Files are placed in your local system i.e., the gateway node**

**You can view the jar files by using the following command**

**ls /data/trendytech/mapreduce_jars**

```
[itv005857@g01 ~]$ ls /data/trendytech/mapreduce_jars/
mapreduce_prog_0_reducer.jar  mapreduce_prog_2_reducer.jar  mapreduce_prog_combiner.jar  mapreduce_prog_cpartitioner.jar  mapreduce_prog.jar
[itv005857@g01 ~]$
```

## Here is how you can run the programs-

**Pre steps before running the Jars**

**-Create input folder named input under your home directory (/user/<username>/data/) in HDFS**

**Ex : hadoop fs -mkdir /user/itv005857/data/input**

**-Create a file named inputfile.txt on your local gateway node and insert some text. Save and Quit**

**Ex : vi inputfile.txt**

**-Copy the file inputfile.txt to the input directory (created in the first step) in HDFS**

**Ex: hadoop fs -put inputfile.txt /user/itv005857/data/input**

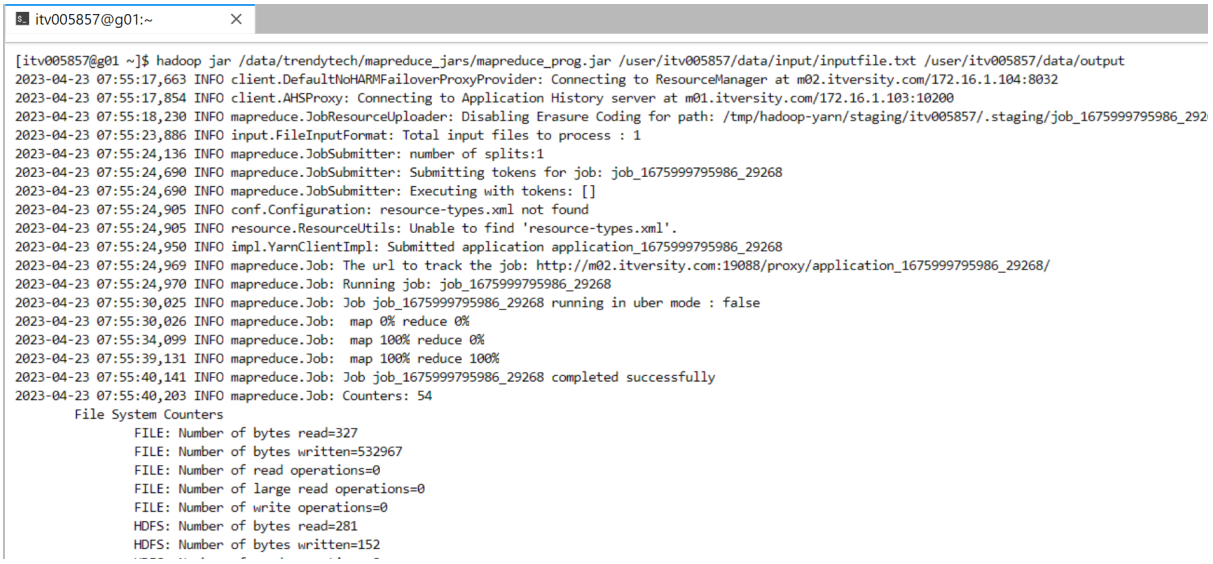
**(The username itv005857 has to be replaced with your respective usernames)**

**Now, you can run the following command -**

**hadoop jar /data/trendytech/mapreduce_jars/<jarname> <input_file_path> <output_folder>**

**(Note - Replace <jarname> <input_file_path> and <output_folder> with their respective values for each of the example programs to be executed)**

**Ex: hadoop jar /data/trendytech/mapreduce_jars/mapreduce_prog.jar /user/itv005857/data/input/inputfile.txt /user/itv005857/data/output**




# Here is boiler plate code to create Spark Session-

```
from pyspark.sql import SparkSession
import getpass
username = getpass.getuser()
spark = SparkSession. \
builder. \
config('spark.ui.port','0'). \
config("spark.sql.warehouse.dir", f"/user/{username}/warehouse"). \
enableHiveSupport(). \
master('yarn'). \
getOrCreate()
```