

# Instructions for Pollution Data Analysis-I

CS555: Big Data Computing Lab

---

**8 September 2020**

## Instructions for Execution

1. Install maven
  - a. Open the terminal and type:  
  
`sudo apt-get install maven`
2. Create Project directory: (Pollution)
  - a. Browse to your home directory.
  - b. Open terminal and type:  
  
`cd /home/iitp`
  - c. Create directory pollution inside your hadoop home directory through terminal  
  
`mkdir Pollution`
  - d. Go into your Pollution directory  
  
`cd Pollution`
3. Create Directory Total
  - a. Browse to Pollution directory, if you are in some other directory  
  
`cd /home/iitp/Pollution`
  - b. Create directory Total  
  
`mkdir Total`
  - c. Browse inside Total directory  
  
`cd Total`

- 
- d. Create the data file into the Total directory

`nano sensorreading.txt`

- e. Copy contents from the source file on your local system,

`Use CTRL-O to save, Press Enter again,And CTRL-X to exit`

- f. Create the pom.xml file

`nano pom.xml`

- g. Repeat the step (e) again

- h. Create directory src

`mkdir src`

- a. Browse into src folder

`cd src`

- b. Create directory main

`mkdir main`

- c. Browse into main folder

`cd main`

- d. Create directory java

`mkdir java`

- e. Browse into java folder

`cd java`

- i. create the source files

- a. Create TotalDriver.java

`nano TotalDriver.java`

- b. Copy contents from the source file on your local system. Use CTRL-O to save, Press Enter again,And CTRL-X to exit

---

c. Create TotalMapper.java

`nano TotalMapper.java`

d. Repeat step (i.b) again

j. Browse to the Total Directory

`cd /home/iitp/Pollution/Total`

k. Compile the java files

`mvn clean && mvn compile && mvn package`

l. Browse to the hadoop directory

`cd /home/iitp/hadoop-2.6.0`

m. create directory Count on the HDFS directory

`bin/hadoop dfs -mkdir /Count`

n. Put the input csv file on to HDFS directory

`bin/hadoop dfs -put /home/iitp/Pollution/Total/sensorreading.txt /Count/`

o. Run the jar file

`bin/hadoop jar /home/iitp/Pollution/Total/target/GVexample-1.0.0.jar TotalDriver  
/Count/ /Count/output`

p. Check the contents of the part file

`bin/hadoop dfs -cat /Count/output/part-r-00000`