

# DSBDAL Practical Assignment

**Assignment Group A-2 :-** Design a distributed application using MapReduce(Using Java) which processes a log file of a system. List out the users who have logged for maximum period on the system. Use simple log file from the Internet and process it using a pseudo distribution mode on Hadoop platform.

- **IPcount on Hadoop:-**

1. Open Eclipse
2. Create Java Project  
Project Name-IPcount
3. Right Click on IPcount Project and create package Name.
4. Right Click on IPcount and Create Class File  
Name+Include main function
5. Add Java Project Code.(separate java code file is given)

- **Remove Error in Code:-**

- 1) Change Class Name-IPcount (Project Name)

**For Library Error**

- 2) Right Click on Project → Click Build Path → Add External  
Archives → User → Library → Hadoop → Hadoop.common.2.6.jar

**Normal Error**

- 3) Right Click on Project → Click Build Path → Add External  
Archives → User → Library → Hadoop2.0 mapreduce → Hadoop.core.2.6.jar

- **Create Jar File:-**

1) Right Click on Project → Export → Java → jar → Browse → Name.jar

- **Open Terminal:-**

Copy log file from Desktop to Present Working Directory(PWD)

→hadoop fs -put Desktop/log\_file.txt input.txt

→hadoop fs -ls

- **Execute /Run .jar file-**

- Hadoop Process Started

→hadoop jar IPcount.jar IPcount.IPcount input.txt output

**Concat One File-**

→hadoop fs -cat output/part-r-00000

- **Copy file From host to (Virtual Machine)VM-**

→Su

→mount -t vboxsf Source\_Machine\_Folder→Destination\_Machine\_Folder

or simply drag and drop file