

### **How to run the the Hadoop Program**

- 1) Upload input\_dataset file to Hadoop file system using below command  
hdfs dfs -put /local path of input file /destination path in hadoop file system  
Ex. Hdfs dfs -put /wordcount\_dataset.txt /input  
Here "input" directory is present in / directory of hadoop file system
- 2) Open a terminal where jar file of the application is present. The jar file is generated after building the program.

#### **Location of jar file :**

- i) WordCount/dist/
  - ii) Summation/dist/
- 3) To run the program, execute the below command on the terminal  
hadoop jar /location of jar file in local filesystem /location of input data set in hadoop file system /location of output directory  
ex. hadoop jar ./WordCount.jar /input/wordcount\_dataset.txt /out1  
here "out1" is a new directory that will contain the result. The result will be collected in a file name "part-00000"  
Note: always give a new name of directory. The directory should not exist earlier in hadoop file system.
  - 4) Download the result in local file system  
hdfs dfs -get /out1/part-00000 ./

### **Compilation of Hadoop Program**

- 1) Open Netbeans IDE or Eclipse
- 2) Write a program
- 3) Export all the jar file needed for the program
- 4) Build the program. To build the program click on "clean and build project" item in "Run" Menu. This will create a "dist" folder in your project directory that contains jar file of the program.