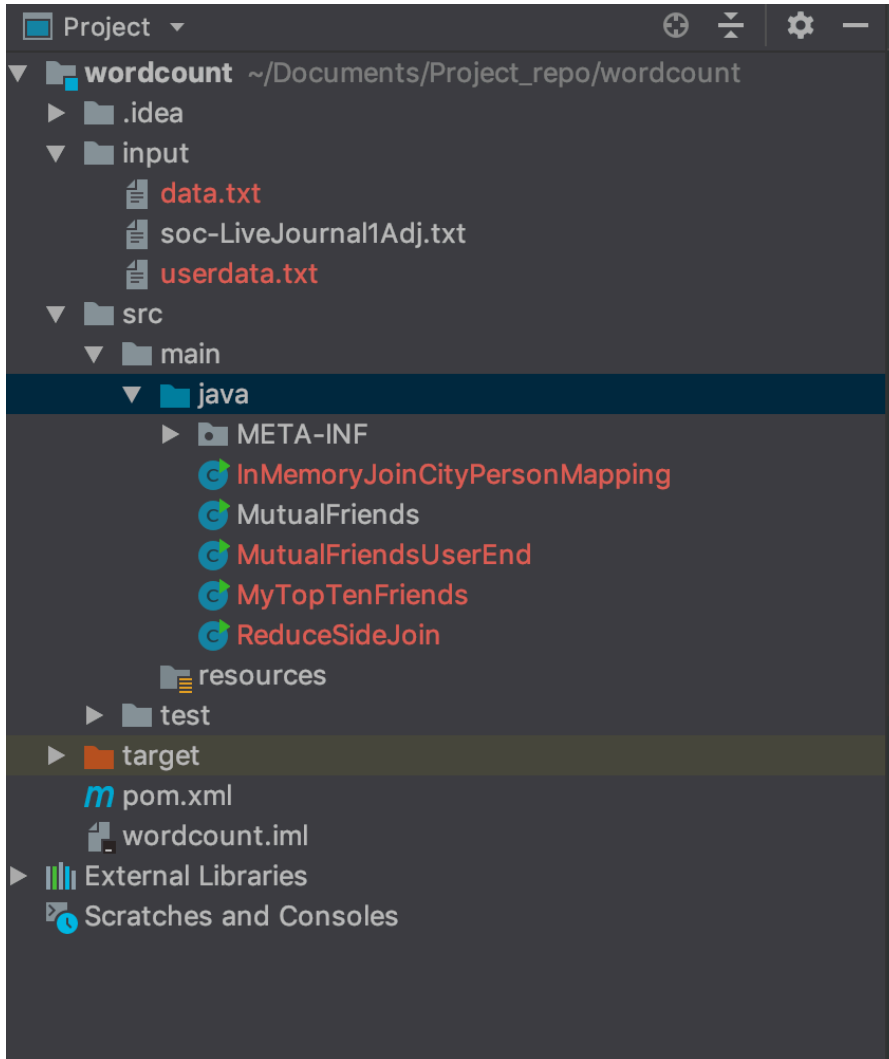


# Friends Data Optimizer Big Data Hadoop

## Project details:

Jar file: wordcount.jar

The package contains project directory as specified in screenshot



Question1 : MutualFriends.java

Question 2 : MyTopTenFriends.java

Question 3 : In MemoryJoin.java

Question 4 : ReduceSideJoin.java

Create dir in hdfs and put data files in it before executing the jar

Use command :

```
hadoop fs -mkdir /user/
```

```
hadoop fs -mkdir /user/harshverma/
```

Put data from local directory to Hadoop system :

```
hadoop fs -put ~/Documents/Project_repo/wordcount/input/userdata.txt /user/harshverma
```

Each question has its own java file within the above-mentioned jar file.

### 1. Filename: MutualFriends.java

Command to run the program:

```
Hadoop jar wordcount.jar MutualFriends /home/harsh/soc-LiveJournal1Adj.txt output1
```

To view the output: `hadoop fs -cat /user/harshverma/output1/part-r-00000`

**Output For the following given pairs:**

```
(0,1) : 5,20
(20, 28193) : 1
(1, 29826)
(6222, 19272) : 19263, 19280, 19281, 19282
(28041, 28056) : 6245, 28054, 28061
```

### 2. Filename: MyTopTenFriends.java

Generic command to run:

```
hadoop jar wordcount.jar MyTopTenFriends <userA> <userB> /home/harsh/soc-LiveJournal1Adj.txt output_1
```

Command to run(with example):

```
hadoop jar Assignment1.jar MyTopTenFriends 0 1 /home/harsh/soc-LiveJournal1Adj.txt output_1
```

To view the output: `hadoop fs -cat /user/harshverma/output1/part-r-00000`

[illegible]

Generic command to run:

```
hadoop jar wordcount.jar InMemoryJoinCityPersonMapping 0 1
/home/harsh/soc-LiveJournal1Adj.txt intmd_output
/home/harsh/userdata.txt output_2
```

0,41 [Harry:Indianola,Robert:Tampa]

To view the output: `hdfs dfs -cat output_3/*`

## Below is the output snapshot:

Virginia Virginia,80,3645 Boggess Street,Dayton,Ohio,22.0  
Santos Santos,79,4496 Fairfield Road,Racine,Wisconsin,22.0  
Richard Richard,69,1905 Martha Ellen Drive,Pahrump,Nevada,22.0  
Jane Jane,54,2824 Clifford Street,San Leandro,California,22.0  
Jim Jim,23,3848 Heather Sees Way,Tulsa,Oklahoma,22.0  
Jackie Jackie,66,221 Amethyst Drive,Lansing,Michigan,22.0  
Arthur Arthur,65,3326 Stratford Drive,Honolulu,Hawaii,22.0  
Sarah Sarah,85,1999 Bagwell Avenue,Gainesville,Florida,22.0  
Elizabeth Elizabeth,24,2846 Walnut Hill Drive,Sharonville,Ohio,22.0  
Audrey Audrey,43,1780 Russell Street,Lawrence,Massachusetts,22.0  
Ramon Ramon,68,3146 Indiana Avenue,Wahiawa,Hawaii,22.0  
Felix Felix,70,2616 Star Trek Drive,Fort Walton Beach,Florida,22.0  
Sha Sha,86,3658 Maple Lane,Gadsden,Alabama,22.0  
Amber Amber,22,461 Pringle Drive,Northbrook,Illinois,22.0  
Ernest Ernest,62,574 Formula Lane,Mc Kinney,Texas,22.0

```
-rw-r--r-- 1 root supergroup      888 2019-03-05 16:57 /user/harshverma/output10/part-r-00000
➔ output git:(master) * hadoop fs -cat /user/harshverma/output10/part-r-00000
2019-03-05 16:57:47,706 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Virginia Virginia,80,3645 Boggess Street,Dayton,Ohio,22.0
Santos Santos,79,4496 Fairfield Road,Racine,Wisconsin,22.0
Richard Richard,69,1905 Martha Ellen Drive,Pahrump,Nevada,22.0
Jane Jane,54,2824 Clifford Street,San Leandro,California,22.0
Jim Jim,23,3848 Heather Sees Way,Tulsa,Oklahoma,22.0
Jackie Jackie,66,221 Amethyst Drive,Lansing,Michigan,22.0
Arthur Arthur,65,3326 Stratford Drive,Honolulu,Hawaii,22.0
Sarah Sarah,85,1999 Bagwell Avenue,Gainesville,Florida,22.0
Elizabeth Elizabeth,24,2846 Walnut Hill Drive,Sharonville,Ohio,22.0
Audrey Audrey,43,1780 Russell Street,Lawrence,Massachusetts,22.0
Ramon Ramon,68,3146 Indiana Avenue,Wahiawa,Hawaii,22.0
Felix Felix,70,2616 Star Trek Drive,Fort Walton Beach,Florida,22.0
Sha Sha,86,3658 Maple Lane,Gadsden,Alabama,22.0
Amber Amber,22,461 Pringle Drive,Northbrook,Illinois,22.0
Ernest Ernest,62,574 Formula Lane,Mc Kinney,Texas,22.0
```