

Mutual Friends

The platform used to execute the assignment is UBUNTU 18.04 and the language used is python.

Project directory : Mutual_Friends

The Mutual_Friends contains following files :

- **ip.txt** : input file for map reduce.
- **mapper.py** : mapper code for the map reduce.
- **reducer.py** : reducer code for the map reduce.
- **hadoop-streaming-2.7.3.jar** : this file is used for streaming utility which is needed in running the map reduce program in python.
- **Input and Output snaps**

Running the Program :

- Create a Project directory in hadoop using :

```
$ hadoop fs -mkdir /Mutual_Friends
```

- Create a Input directory in project Mutual_Friends

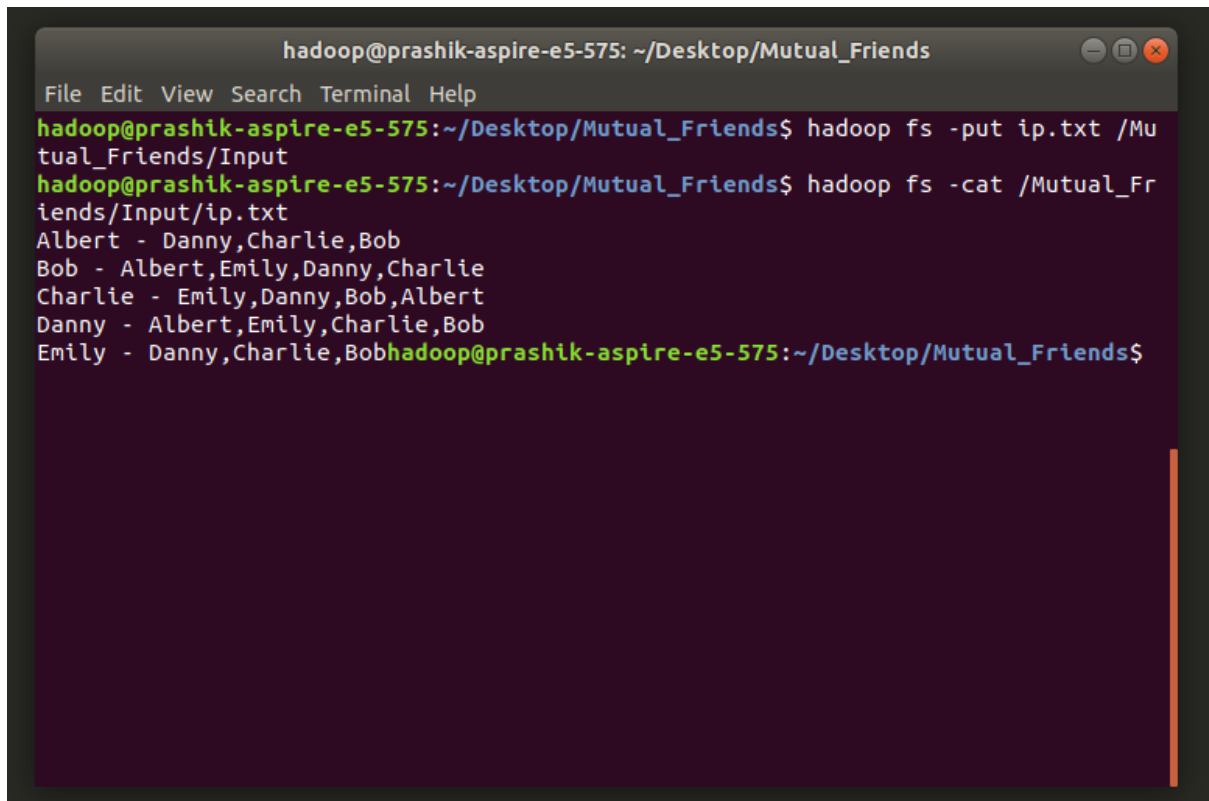
```
$ hadoop fs -mkdir /Mutual_Friends/Input
```

- Put your input file (ip.txt) in HDFS using :

```
$ hadoop fs -put ip.txt /Mutual_Friends/Input
```

- View the input file in HDFS using command :

```
$ hadoop fs -cat /Mutual_Friends/Input/ip.txt
```

A terminal window titled 'hadoop@prashik-aspire-e5-575: ~/Desktop/Mutual_Friends'. The window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The terminal shows the following commands and output:

```
hadoop@prashik-aspire-e5-575:~/Desktop/Mutual_Friends$ hadoop fs -put ip.txt /Mutual_Friends/Input
hadoop@prashik-aspire-e5-575:~/Desktop/Mutual_Friends$ hadoop fs -cat /Mutual_Friends/Input/ip.txt
Albert - Danny,Charlie,Bob
Bob - Albert,Emily,Danny,Charlie
Charlie - Emily,Danny,Bob,Albert
Danny - Albert,Emily,Charlie,Bob
Emily - Danny,Charlie,Bob
```

- Run the map reduce program using command :

```
$ hadoop jar <hadoop-streaming_path> -input <hdfs_input_path> -output  
<hdfs_output_path> -mapper <mapper_code_path> -reducer  
<reducer_code_path>
```

For example :

```
$ hadoop jar '/home/hadoop/Desktop/Mutual_Friends/hadoop-streaming-  
2.7.3.jar' -input /Mutual_Friends/input/ip.txt -output  
/Mutual_Friends/output -mapper  
'/home/hadoop/Desktop/Mutual_Friends/mapper.py' -reducer  
'/home/hadoop/Desktop/Mutual_Friends/reducer.py'
```

```
Activities Terminal Sun 21:55
hadoop@prashik-aspire-e5-575: ~/Desktop/Mutual_Friends

hadoop@prashik-aspire-e5-575:~/Desktop/Mutual_Friends$ hadoop jar '/home/hadoop/Desktop/Mutual_Friends/hadoop-streaming-2.7.3.jar' -input /Mutual_Friends/Input/ip.txt -output /Mutual_Friends/Output -mapper '/home/hadoop/Desktop/Mutual_Friends/mapper.py' -reducer '/home/hadoop/Desktop/Mutual_Friends/reducer.py'
packageJobJar: [/tmp/hadoop-unjar7606519440415406495/] [] /tmp/streamjob7526185339925350568.jar tmpDir=null
20/04/26 21:54:29 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
20/04/26 21:54:29 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
20/04/26 21:54:30 INFO mapred.FileInputFormat: Total input files to process : 1
20/04/26 21:54:30 INFO mapreduce.JobSubmitter: number of splits:2
20/04/26 21:54:31 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1587852006321_0085
20/04/26 21:54:31 INFO impl.YarnClientImpl: Submitted application application_1587852006321_0085
20/04/26 21:54:31 INFO mapreduce.Job: The url to track the job: http://prashik-Aspire-E5-575:8088/proxy/application_1587852006321_0085/
20/04/26 21:54:31 INFO mapreduce.Job: Running job: job_1587852006321_0085
20/04/26 21:54:49 INFO mapreduce.Job: Job job_1587852006321_0085 running in uber mode : false
20/04/26 21:54:49 INFO mapreduce.Job: map 0% reduce 0%
20/04/26 21:54:57 INFO mapreduce.Job: map 50% reduce 0%
20/04/26 21:54:58 INFO mapreduce.Job: map 100% reduce 0%
20/04/26 21:55:04 INFO mapreduce.Job: map 100% reduce 100%
20/04/26 21:55:05 INFO mapreduce.Job: Job job_1587852006321_0085 completed successfully
20/04/26 21:55:05 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=672
    FILE: Number of bytes written=481475
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=429
    HDFS: Number of bytes written=564
    HDFS: Number of read operations=9
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Launched map tasks=2
    Launched reduce tasks=1
    Data-local map tasks=2
    Total time spent by all maps in occupied slots (ms)=12746
    Total time spent by all reduces in occupied slots (ms)=4141
    Total time spent by all map tasks (ms)=12746
    Total time spent by all reduce tasks (ms)=4141
```

- View the map reduce output using command :

\$ hadoop fs -cat /Mutual_Friends/output/part-00000

```
Activities Terminal Sun 21:56
hadoop@prashik-aspire-e5-575: ~/Desktop/Mutual_Friends

Merged Map outputs=2
GC time elapsed (ms)=550
CPU time spent (ms)=4600
Physical memory (bytes) snapshot=723591168
Virtual memory (bytes) snapshot=5899337728
Total committed heap usage (bytes)=519045120
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=227
File Output Format Counters
  Bytes Written=564
20/04/26 21:55:05 INFO streaming.StreamJob: Output directory: /Mutual_Friends/Output
hadoop@prashik-aspire-e5-575:~/Desktop/Mutual_Friends$ hadoop fs -cat /Mutual_Friends/Output/part-00000
Mutual Friends of : Albert,Bob ----->>> Charlie,Danny
Mutual Friends of : Albert,Charlie ----->>> Bob,Danny
Mutual Friends of : Albert,Danny ----->>> Bob,Charlie
Mutual Friends of : Bob,Charlie ----->>> Albert,Danny,Emily
Mutual Friends of : Bob,Danny ----->>> Albert,Charlie,Emily
Mutual Friends of : Bob,Emily ----->>> Charlie,Danny
Mutual Friends of : Charlie,Danny ----->>> Albert,Bob,Emily
Mutual Friends of : Charlie,Emily ----->>> Bob,Danny
Mutual Friends of : Danny,Emily ----->>> Bob,Charlie
hadoop@prashik-aspire-e5-575:~/Desktop/Mutual_Friends$
```

