


```

20/11/26 21:22:51 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement
the Tool interface and execute your application with ToolRunner to remedy this.
20/11/26 21:22:52 INFO input.FileInputFormat: Total input paths to process : 1
20/11/26 21:22:52 INFO mapreduce.JobSubmitter: number of splits:1
20/11/26 21:22:52 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1604346392376_4833
20/11/26 21:22:53 INFO impl.YarnClientImpl: Submitted application application_1604346392376_4833
20/11/26 21:22:53 INFO mapreduce.Job: The url to track the job: http://babar.es.its.nyu.edu:8088/proxy/application
_1604346392376_4833/
20/11/26 21:22:53 INFO mapreduce.Job: Running job: job_1604346392376_4833
20/11/26 21:22:57 INFO mapreduce.Job: Job job_1604346392376_4833 running in uber mode : false
20/11/26 21:22:57 INFO mapreduce.Job: map 0% reduce 0%
20/11/26 21:23:03 INFO mapreduce.Job: map 100% reduce 0%
20/11/26 21:23:08 INFO mapreduce.Job: map 100% reduce 78%
20/11/26 21:23:09 INFO mapreduce.Job: map 100% reduce 100%
20/11/26 21:23:11 INFO mapreduce.Job: Job job_1604346392376_4833 completed successfully
20/11/26 21:23:11 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=1268854
    FILE: Number of bytes written=7625219
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=41156268
    HDFS: Number of bytes written=1816
    HDFS: Number of read operations=99
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=64
  Job Counters
    Launched map tasks=1
    Launched reduce tasks=32
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=15228
    Total time spent by all reduces in occupied slots (ms)=524640
    Total time spent by all map tasks (ms)=3807
    Total time spent by all reduce tasks (ms)=87440
    Total vcore-milliseconds taken by all map tasks=3807
    Total vcore-milliseconds taken by all reduce tasks=87440
    Total megabyte-milliseconds taken by all map tasks=15593472
    Total megabyte-milliseconds taken by all reduce tasks=537231360
  Map-Reduce Framework
    Map input records=1048576
    Map output records=1048576
    Map output bytes=24375447
    Map output materialized bytes=1268726
    Input split bytes=111
    Combine input records=0
    Combine output records=0
    Reduce input groups=70
    Reduce shuffle bytes=1268726
    Reduce input records=1048576
    Reduce output records=70
    Spilled Records=2097152
    Shuffled Maps =32
    Failed Shuffles=0
    Merged Map outputs=32
    GC time elapsed (ms)=1686
    CPU time spent (ms)=59460
    Physical memory (bytes) snapshot=12922327040
    Virtual memory (bytes) snapshot=123582869504
    Total committed heap usage (bytes)=39885209600
  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=41156157
  File Output Format Counters
    Bytes Written=1816
[lh2514@login-2-1 ~]$ hdfs dfs -ls /user/lh2514/class1/NYC_output35
Found 33 items
-rw-r--r--+ 3 lh2514 users          0 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/_SUCCESS
-rw-r--r--+ 3 lh2514 users        35 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00000

```

```

-rw-r--r--+ 3 lh2514 users 30 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00001
-rw-r--r--+ 3 lh2514 users 35 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00002
-rw-r--r--+ 3 lh2514 users 37 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00003
-rw-r--r--+ 3 lh2514 users 55 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00004
-rw-r--r--+ 3 lh2514 users 118 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00005
-rw-r--r--+ 3 lh2514 users 25 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00006
-rw-r--r--+ 3 lh2514 users 36 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00007
-rw-r--r--+ 3 lh2514 users 65 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00008
-rw-r--r--+ 3 lh2514 users 36 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00009
-rw-r--r--+ 3 lh2514 users 46 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00010
-rw-r--r--+ 3 lh2514 users 21 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00011
-rw-r--r--+ 3 lh2514 users 29 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00012
-rw-r--r--+ 3 lh2514 users 99 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00013
-rw-r--r--+ 3 lh2514 users 77 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00014
-rw-r--r--+ 3 lh2514 users 57 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00015
-rw-r--r--+ 3 lh2514 users 44 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00016
-rw-r--r--+ 3 lh2514 users 38 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00017
-rw-r--r--+ 3 lh2514 users 54 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00018
-rw-r--r--+ 3 lh2514 users 63 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00019
-rw-r--r--+ 3 lh2514 users 69 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00020
-rw-r--r--+ 3 lh2514 users 73 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00021
-rw-r--r--+ 3 lh2514 users 0 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00022
-rw-r--r--+ 3 lh2514 users 90 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00023
-rw-r--r--+ 3 lh2514 users 0 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00024
-rw-r--r--+ 3 lh2514 users 137 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00025
-rw-r--r--+ 3 lh2514 users 0 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00026
-rw-r--r--+ 3 lh2514 users 117 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00027
-rw-r--r--+ 3 lh2514 users 154 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00028
-rw-r--r--+ 3 lh2514 users 82 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00029
-rw-r--r--+ 3 lh2514 users 58 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00030
-rw-r--r--+ 3 lh2514 users 36 2020-11-26 21:23 /user/lh2514/class1/NYC_output35/part-r-00031

[lh2514@login-2-1 ~]$ vi CrimesCountRunner.java
[lh2514@login-2-1 ~]$ javac -classpath "$(yarn classpath)":.. -d . CrimesCountRunner.java
[lh2514@login-2-1 ~]$ jar -cvf CrimeCount.jar *.class
added manifest
adding: Airbnb.class(in = 1428) (out= 789)(deflated 44%)
adding: AirbnbMapper.class(in = 2269) (out= 1047)(deflated 53%)
adding: AirbnbReducer.class(in = 1694) (out= 714)(deflated 57%)
adding: CrimeCountMapper.class(in = 1743) (out= 735)(deflated 57%)
adding: CrimesCountRunner.class(in = 1416) (out= 777)(deflated 45%)
adding: IntSumReducer.class(in = 1694) (out= 714)(deflated 57%)
adding: LineCountMapper.class(in = 1478) (out= 596)(deflated 59%)
adding: LinesCountRunner.class(in = 1374) (out= 755)(deflated 45%)
adding: Main.class(in = 1461) (out= 872)(deflated 40%)
adding: SelectFieldsMapper.class(in = 255) (out= 198)(deflated 22%)
adding: SelectFieldsReducer.class(in = 258) (out= 196)(deflated 24%)
adding: SelectFieldsRunner.class(in = 1765) (out= 933)(deflated 47%)
adding: WordCount.class(in = 1440) (out= 789)(deflated 45%)
adding: WordCountMapper.class(in = 2096) (out= 962)(deflated 54%)
adding: WordCountReducer.class(in = 1700) (out= 717)(deflated 57%)
[lh2514@login-2-1 ~]$ hadoop jar CrimeCount.jar CrimesCountRunner /user/lh2514/class1/NYCCleaned.csv /user/lh2514/class1/NYC_output38
20/11/26 21:25:59 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
20/11/26 21:26:00 INFO input.FileInputFormat: Total input paths to process : 1
20/11/26 21:26:01 INFO mapreduce.JobSubmitter: number of splits:1
20/11/26 21:26:01 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1604346392376_4834
20/11/26 21:26:02 INFO impl.YarnClientImpl: Submitted application application_1604346392376_4834
20/11/26 21:26:02 INFO mapreduce.Job: The url to track the job: http://babar.es.its.nyu.edu:8088/proxy/application_1604346392376_4834/
20/11/26 21:26:02 INFO mapreduce.Job: Running job: job_1604346392376_4834
20/11/26 21:26:07 INFO mapreduce.Job: Job job_1604346392376_4834 running in uber mode : false
20/11/26 21:26:07 INFO mapreduce.Job: map 0% reduce 0%
20/11/26 21:26:12 INFO mapreduce.Job: map 100% reduce 0%
20/11/26 21:26:18 INFO mapreduce.Job: map 100% reduce 100%
20/11/26 21:26:21 INFO mapreduce.Job: Job job_1604346392376_4834 completed successfully
20/11/26 21:26:21 INFO mapreduce.Job: Counters: 49
File System Counters
FILE: Number of bytes read=1268410
FILE: Number of bytes written=2845161
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=41156268
HDFS: Number of bytes written=1816

```

HDFS: Number of read operations=6
HDFS: Number of large read operations=0
HDFS: Number of write operations=2

Job Counters

Launched map tasks=1
Launched reduce tasks=1
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=14836
Total time spent by all reduces in occupied slots (ms)=19428
Total time spent by all map tasks (ms)=3709
Total time spent by all reduce tasks (ms)=3238
Total vcore-milliseconds taken by all map tasks=3709
Total vcore-milliseconds taken by all reduce tasks=3238
Total megabyte-milliseconds taken by all map tasks=15192064
Total megabyte-milliseconds taken by all reduce tasks=19894272

Map-Reduce Framework

Map input records=1048576
Map output records=1048576
Map output bytes=24375447
Map output materialized bytes=1268406
Input split bytes=111
Combine input records=0
Combine output records=0
Reduce input groups=70
Reduce shuffle bytes=1268406
Reduce input records=1048576
Reduce output records=70
Spilled Records=2097152
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=408
CPU time spent (ms)=11890
Physical memory (bytes) snapshot=1537216512
Virtual memory (bytes) snapshot=7477518336
Total committed heap usage (bytes)=2807037952

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=41156157

File Output Format Counters

Bytes Written=1816

```
[lh2514@login-2-1 ~]$ hdfs dfs -ls /user/lh2514/class1/NYC_output38
```

Found 2 items

```
-rw-r--r--+ 3 lh2514 users          0 2020-11-26 21:26 /user/lh2514/class1/NYC_output38/_SUCCESS
-rw-r--r--+ 3 lh2514 users       1816 2020-11-26 21:26 /user/lh2514/class1/NYC_output38/part-r-00000
```

```
[lh2514@login-2-1 ~]$ hdfs dfs -cat /user/lh2514/class1/NYC_output38/part-r-00000
```

2880

```
"HOMICIDE-NEGLIGENT      8
"LOITERING/GAMBLING (CARDS    19
ABORTION                  2
ADMINISTRATIVE CODE       2335
ADMINISTRATIVE CODES       2
AGRICULTURE & MRKTS LAW-UNCLASSIFIED    83
ALCOHOLIC BEVERAGE CONTROL LAW  161
ANTICIPATORY OFFENSES      9
ARSON                    2472
ASSAULT 3 & RELATED OFFENSES  114430
BURGLAR'S TOOLS          716
BURGLARY                 34994
CHILD ABANDONMENT/NON SUPPORT    61
CRIMINAL MISCHIEF & RELATED OF  102771
CRIMINAL TRESPASS         10292
DANGEROUS DRUGS          56868
DANGEROUS WEAPONS         22953
DISORDERLY CONDUCT        123
DISRUPTION OF A RELIGIOUS SERV  13
ENDAN WELFARE INCOMP       52
ESCAPE 3                  25
FELONY ASSAULT           43921
```

```

FORGERY 10591
FORTUNE TELLING 1
FRAUDS 7861
FRAUDULENT ACCOSTING 455
GAMBLING 545
GRAND LARCENY 96232
GRAND LARCENY OF MOTOR VEHICLE 16223
HARRASSMENT 2 133179
HOMICIDE-NEGLIGENT-VEHICLE 7
INTOXICATED & IMPAIRED DRIVING 15169
INTOXICATED/IMPAIRED DRIVING 18
JOSTLING 80
KIDNAPPING 2
KIDNAPPING & RELATED OFFENSES 332
KIDNAPPING AND RELATED OFFENSES 2
LOITERING 2
LOITERING FOR DRUG PURPOSES 1
MISCELLANEOUS PENAL LAW 29221
MURDER & NON-NEGL. MANSLAUGHTER 733
NEW YORK CITY HEALTH CODE 25
NYS LAWS-UNCLASSIFIED FELONY 958
NYS LAWS-UNCLASSIFIED VIOLATION 6
OFF. AGNST PUB ORD SENSBLY & 44772
OFFENSES AGAINST PUBLIC ADMINI 21353
OFFENSES AGAINST PUBLIC SAFETY 184
OFFENSES AGAINST THE PERSON 2804
OFFENSES INVOLVING FRAUD 2966
OFFENSES RELATED TO CHILDREN 303
OFNS_DESC 1
OTHER OFFENSES RELATED TO THEF 3081
OTHER STATE LAWS 31
OTHER STATE LAWS (NON PENAL LA 1035
OTHER STATE LAWS (NON PENAL LAW) 1
OTHER TRAFFIC INFRACTION 3
PETIT LARCENY 180246
PETIT LARCENY OF MOTOR VEHICLE 156
POSSESSION OF STOLEN PROPERTY 4928
PROSTITUTION & RELATED OFFENSES 240
RAPE 2978
ROBBERY 36801
SEX CRIMES 11780
THEFT OF SERVICES 993
THEFT-FRAUD 10472
UNAUTHORIZED USE OF A VEHICLE 3555
UNDER THE INFLUENCE OF DRUGS 2
UNLAWFUL POSS. WEAP. ON SCHOOL 8
VEHICLE AND TRAFFIC LAWS 13050
[lh2514@login-2-1 ~]$ vi CrimeCountMapper.java
[lh2514@login-2-1 ~]$ javac -classpath "$(yarn classpath)" -d . CrimeCountMapper.java
[lh2514@login-2-1 ~]$ javac -classpath "$(yarn classpath)":. -d . CrimesCountRunner.java
[lh2514@login-2-1 ~]$ jar -cvf CrimeCount.jar *.class
added manifest
adding: Airbnb.class(in = 1428) (out= 789)(deflated 44%)
adding: AirbnbMapper.class(in = 2269) (out= 1047)(deflated 53%)
adding: AirbnbReducer.class(in = 1694) (out= 714)(deflated 57%)
adding: CrimeCountMapper.class(in = 1743) (out= 735)(deflated 57%)
adding: CrimesCountRunner.class(in = 1416) (out= 777)(deflated 45%)
adding: IntSumReducer.class(in = 1694) (out= 714)(deflated 57%)
adding: LineCountMapper.class(in = 1478) (out= 596)(deflated 59%)
adding: LinesCountRunner.class(in = 1374) (out= 755)(deflated 45%)
adding: Main.class(in = 1461) (out= 872)(deflated 40%)
adding: SelectFieldsMapper.class(in = 255) (out= 198)(deflated 22%)
adding: SelectFieldsReducer.class(in = 258) (out= 196)(deflated 24%)
adding: SelectFieldsRunner.class(in = 1765) (out= 933)(deflated 47%)
adding: WordCount.class(in = 1440) (out= 789)(deflated 45%)
adding: WordCountMapper.class(in = 2096) (out= 962)(deflated 54%)
adding: WordCountReducer.class(in = 1700) (out= 717)(deflated 57%)
[lh2514@login-2-1 ~]$ hadoop jar CrimeCount.jar CrimesCountRunner /user/lh2514/class1/NYCcleaned.csv /user/lh2514/class1/NYC_output38
20/11/26 21:29:17 WARN security.UserGroupInformation: PriviledgedActionException as:lh2514 (auth:SIMPLE) cause:org
.apache.hadoop.mapred.FileAlreadyExistsException: Output directory hdfs://dumbo/user/lh2514/class1/NYC_output38 al
ready exists
Exception in thread "main" org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory hdfs://dumbo/user
/lh2514/class1/NYC_output38 already exists
    at org.apache.hadoop.mapreduce.lib.output.FileOutputFormat.checkOutputSpecs(FileOutputFormat.java:146)

```

```

at org.apache.hadoop.mapreduce.JobSubmitter.checkSpecs(JobSubmitter.java:270)
at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:143)
at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1307)
at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1304)
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:422)
at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1924)
at org.apache.hadoop.mapreduce.Job.submit(Job.java:1304)
at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1325)
at CrimesCountRunner.main(CrimesCountRunner.java:23)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.hadoop.util.RunJar.run(RunJar.java:226)
at org.apache.hadoop.util.RunJar.main(RunJar.java:141)
[lh2514@login-2-1 ~]$ hadoop jar CrimeCount.jar CrimesCountRunner /user/lh2514/class1/NYCcleaned.csv /user/lh2514/class1/NYC_output39
20/11/26 21:29:32 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement
the Tool interface and execute your application with ToolRunner to remedy this.
20/11/26 21:29:32 INFO input.FileInputFormat: Total input paths to process : 1
20/11/26 21:29:33 INFO mapreduce.JobSubmitter: number of splits:1
20/11/26 21:29:33 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1604346392376_4836
20/11/26 21:29:34 INFO impl.YarnClientImpl: Submitted application application_1604346392376_4836
20/11/26 21:29:34 INFO mapreduce.Job: The url to track the job: http://babar.es.its.nyu.edu:8088/proxy/application
_1604346392376_4836/
20/11/26 21:29:34 INFO mapreduce.Job: Running job: job_1604346392376_4836
20/11/26 21:29:38 INFO mapreduce.Job: Job job_1604346392376_4836 running in uber mode : false
20/11/26 21:29:38 INFO mapreduce.Job: map 0% reduce 0%
20/11/26 21:29:44 INFO mapreduce.Job: map 100% reduce 0%
20/11/26 21:29:49 INFO mapreduce.Job: map 100% reduce 100%
20/11/26 21:30:03 INFO mapreduce.Job: Job job_1604346392376_4836 completed successfully
20/11/26 21:30:03 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=411051
    FILE: Number of bytes written=1130443
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=41156268
    HDFS: Number of bytes written=120
    HDFS: Number of read operations=6
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=14116
    Total time spent by all reduces in occupied slots (ms)=19044
    Total time spent by all map tasks (ms)=3529
    Total time spent by all reduce tasks (ms)=3174
    Total vcore-milliseconds taken by all map tasks=3529
    Total vcore-milliseconds taken by all reduce tasks=3174
    Total megabyte-milliseconds taken by all map tasks=14454784
    Total megabyte-milliseconds taken by all reduce tasks=19501056
  Map-Reduce Framework
    Map input records=1048576
    Map output records=1048576
    Map output bytes=6604213
    Map output materialized bytes=411047
    Input split bytes=111
    Combine input records=0
    Combine output records=0
    Reduce input groups=14
    Reduce shuffle bytes=411047
    Reduce input records=1048576
    Reduce output records=14
    Spilled Records=2097152
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=411
    CPU time spent (ms)=11260
    Physical memory (bytes) snapshot=1500246016

```

```

Virtual memory (bytes) snapshot=7474839552
Total committed heap usage (bytes)=2818048000
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=41156157
File Output Format Counters
  Bytes Written=120
[lh2514@login-2-1 ~]$ hdfs dfs -cat /user/lh2514/class1/NYC_output39/part-r-00000
65
1      74778
10     87570
11     114936
12     110305
2      65837
3      78740
4      79373
5      87352
6      84739
7      89611
8      90052
9      85217
CMPLNT_FR_DT      1
[lh2514@login-2-1 ~]$ vi CrimeCount.Mapper
[lh2514@login-2-1 ~]$ vi CrimeCountMapper.java
[lh2514@login-2-1 ~]$ javac -classpath "$(yarn classpath)" -d . CrimeCountMapper.java
[lh2514@login-2-1 ~]$ javac -classpath "$(yarn classpath)":. -d . CrimesCountRunner.java
[lh2514@login-2-1 ~]$ jar -cvf CrimeCount.jar *.class
added manifest
adding: Airbnb.class(in = 1428) (out= 789)(deflated 44%)
adding: AirbnbMapper.class(in = 2269) (out= 1047)(deflated 53%)
adding: AirbnbReducer.class(in = 1694) (out= 714)(deflated 57%)
adding: CrimeCountMapper.class(in = 1743) (out= 735)(deflated 57%)
adding: CrimesCountRunner.class(in = 1416) (out= 777)(deflated 45%)
adding: IntSumReducer.class(in = 1694) (out= 714)(deflated 57%)
adding: LineCountMapper.class(in = 1478) (out= 596)(deflated 59%)
adding: LinesCountRunner.class(in = 1374) (out= 755)(deflated 45%)
adding: Main.class(in = 1461) (out= 872)(deflated 40%)
adding: SelectFieldsMapper.class(in = 255) (out= 198)(deflated 22%)
adding: SelectFieldsReducer.class(in = 258) (out= 196)(deflated 24%)
adding: SelectFieldsRunner.class(in = 1765) (out= 933)(deflated 47%)
adding: WordCount.class(in = 1440) (out= 789)(deflated 45%)
adding: WordCountMapper.class(in = 2096) (out= 962)(deflated 54%)
adding: WordCountReducer.class(in = 1700) (out= 717)(deflated 57%)
[lh2514@login-2-1 ~]$ hadoop jar CrimeCount.jar CrimesCountRunner /user/lh2514/class1/NYCcleaned.csv /user/lh2514/class1/NYC_output40
20/11/26 21:31:48 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement
the Tool interface and execute your application with ToolRunner to remedy this.
20/11/26 21:31:48 INFO input.FileInputFormat: Total input paths to process : 1
20/11/26 21:31:50 INFO mapreduce.JobSubmitter: number of splits:1
20/11/26 21:31:51 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1604346392376_4837
20/11/26 21:31:51 INFO impl.YarnClientImpl: Submitted application application_1604346392376_4837
20/11/26 21:31:51 INFO mapreduce.Job: The url to track the job: http://babar.es.its.nyu.edu:8088/proxy/application
_1604346392376_4837/
20/11/26 21:31:51 INFO mapreduce.Job: Running job: job_1604346392376_4837
20/11/26 21:31:55 INFO mapreduce.Job: Job job_1604346392376_4837 running in uber mode : false
20/11/26 21:31:55 INFO mapreduce.Job: map 0% reduce 0%
20/11/26 21:32:01 INFO mapreduce.Job: map 100% reduce 0%
20/11/26 21:32:06 INFO mapreduce.Job: map 100% reduce 100%
20/11/26 21:32:07 INFO mapreduce.Job: Job job_1604346392376_4837 completed successfully
20/11/26 21:32:07 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=554386
  FILE: Number of bytes written=1417113
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=41156268
  HDFS: Number of bytes written=283
  HDFS: Number of read operations=6

```

HDFS: Number of large read operations=0
HDFS: Number of write operations=2

Job Counters

Launched map tasks=1
Launched reduce tasks=1
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=14956
Total time spent by all reduces in occupied slots (ms)=17388
Total time spent by all map tasks (ms)=3739
Total time spent by all reduce tasks (ms)=2898
Total vcore-milliseconds taken by all map tasks=3739
Total vcore-milliseconds taken by all reduce tasks=2898
Total megabyte-milliseconds taken by all map tasks=15314944
Total megabyte-milliseconds taken by all reduce tasks=17805312

Map-Reduce Framework

Map input records=1048576
Map output records=1048576
Map output bytes=9614583
Map output materialized bytes=554382
Input split bytes=111
Combine input records=0
Combine output records=0
Reduce input groups=25
Reduce shuffle bytes=554382
Reduce input records=1048576
Reduce output records=25
Spilled Records=2097152
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=397
CPU time spent (ms)=11140
Physical memory (bytes) snapshot=1521635328
Virtual memory (bytes) snapshot=7485984768
Total committed heap usage (bytes)=2828009472

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=41156157

File Output Format Counters

Bytes Written=283

[lh2514@login-2-1 ~]\$ hdfs dfs -cat /user/lh2514/class1/NYC_output40/part-r-00000

0 AM 49875
0 PM 58017
1 AM 35901
1 PM 49550
10 AM 40103
10 PM 50302
11 AM 41351
11 PM 45635
2 AM 29934
2 PM 56111
3 AM 25246
3 PM 62273
4 AM 21995
4 PM 60321
5 AM 15659
5 PM 61047
6 AM 15702
6 PM 62200
7 AM 22496
7 PM 59530
8 AM 34923
8 PM 58779
9 AM 38377
9 PM 53248

CMPLNT_FR_TM 1

[lh2514@login-2-1 ~]\$