

```
Last login: Sat Nov 14 19:01:54 on ttys000
(base) lejiahu@Lejias-MBP ~ % ssh -Y lh2514@dumbo.es.its.nyu.edu
Password:
Last login: Sat Nov 14 19:02:34 2020 from a-10-27-14-149.dynapool.vpn.nyu.edu
Rocks 6.2 (SideWinder)
Profile built 16:54 24-Aug-2015
```

```

~~~~~
THIS NYU-OPERATED FACILITY MAY BE ACCESSED ONLY BY AUTHORIZED USERS.
      UNAUTHORIZED PERSONS ... TYPE 'exit' NOW
See https://wikis.nyu.edu/display/NYUHPC/NYU+HPC+Access+Policy for details
of the access policy. Questions regarding the access policy should be
directed (by e-mail) to security@nyu.edu or (by phone) to 212-998-3333
~~~~~

```

```

*****

[lh2514@login-2-1 ~]$ javac -classpath "$(yarn classpath)":.. -d . SelectFieldsRu
nner.java
Note: SelectFieldsRunner.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
[lh2514@login-2-1 ~]$ javac SelectFieldsRunner.java -Xlint:deprecation SelectFie
ldsRunner.java
SelectFieldsRunner.java:1: error: package org.apache.hadoop.conf does not exist
import org.apache.hadoop.conf.Configuration;
                             ^
SelectFieldsRunner.java:2: error: package org.apache.hadoop.fs does not exist
import org.apache.hadoop.fs.Path;
                             ^
SelectFieldsRunner.java:3: error: package org.apache.hadoop.io does not exist
import org.apache.hadoop.io.Text;
                             ^
SelectFieldsRunner.java:4: error: package org.apache.hadoop.mapreduce does not e
xist
import org.apache.hadoop.mapreduce.Job;

```

```

^
SelectFieldsRunner.java:5: error: package org.apache.hadoop.mapreduce.lib.fields
el does not exist
import org.apache.hadoop.mapreduce.lib.fieldsel.FieldSelectionHelper;
^
SelectFieldsRunner.java:6: error: package org.apache.hadoop.mapreduce.lib.input
does not exist
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
^
SelectFieldsRunner.java:7: error: package org.apache.hadoop.mapreduce.lib.output
does not exist
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
^
SelectFieldsRunner.java:17: error: cannot find symbol
    Configuration conf = new Configuration();
    ^
symbol:   class Configuration
location: class SelectFieldsRunner
SelectFieldsRunner.java:17: error: cannot find symbol
    Configuration conf = new Configuration();
    ^
symbol:   class Configuration
location: class SelectFieldsRunner
SelectFieldsRunner.java:19: error: cannot find symbol
    conf.set(FieldSelectionHelper.MAP_OUTPUT_KEY_VALUE_SPEC, ":1,2,7,13");
    ^
symbol:   variable FieldSelectionHelper
location: class SelectFieldsRunner
SelectFieldsRunner.java:20: error: cannot find symbol
    conf.set(FieldSelectionHelper.REDUCE_OUTPUT_KEY_VALUE_SPEC, "0,1,2,3");
    ^
symbol:   variable FieldSelectionHelper
location: class SelectFieldsRunner
SelectFieldsRunner.java:21: error: cannot find symbol
    conf.set(FieldSelectionHelper.DATA_FIELD_SEPERATOR, ",");
    ^
symbol:   variable FieldSelectionHelper
location: class SelectFieldsRunner
SelectFieldsRunner.java:25: error: cannot find symbol
    Job job = new Job(conf, "GroupTest");
    ^
symbol:   class Job
location: class SelectFieldsRunner
SelectFieldsRunner.java:25: error: cannot find symbol
    Job job = new Job(conf, "GroupTest");
    ^
symbol:   class Job
location: class SelectFieldsRunner
SelectFieldsRunner.java:30: error: cannot find symbol
    job.setMapOutputKeyClass(Text.class);
    ^
symbol:   class Text

```

```

location: class SelectFieldsRunner
SelectFieldsRunner.java:31: error: cannot find symbol
    job.setMapOutputValueClass(Text.class);
                                ^
symbol:   class Text
location: class SelectFieldsRunner
SelectFieldsRunner.java:42: error: cannot find symbol
    FileInputFormat.addInputPath(job, new Path("./NYPD_Complaint_Data_Historic.csv"));
                                ^
symbol:   class Path
location: class SelectFieldsRunner
SelectFieldsRunner.java:42: error: cannot find symbol
    FileInputFormat.addInputPath(job, new Path("./NYPD_Complaint_Data_Historic.csv"));
                                ^
symbol:   variable FileInputFormat
location: class SelectFieldsRunner
SelectFieldsRunner.java:43: error: cannot find symbol
    FileOutputFormat.setOutputPath(job, new Path("./fieldsel"));
                                ^
symbol:   class Path
location: class SelectFieldsRunner
SelectFieldsRunner.java:43: error: cannot find symbol
    FileOutputFormat.setOutputPath(job, new Path("./fieldsel"));
                                ^
symbol:   variable FileOutputFormat
location: class SelectFieldsRunner

```

20 errors

```
[lh2514@login-2-1 ~]$ jar -cvf SelectFields.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: 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 = 1819) (out= 964)(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 SelectFields.jar SelectFieldsRunner /user/lh2514/class1/NYPD_Complaint_Data_Historic.csv /user/lh2514/class1/NYC_output1
```

```
20/11/15 08:40:05 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
```

```
20/11/15 08:40:06 INFO mapreduce.JobSubmitter: Cleaning up the staging area /user/lh2514/.staging/job_1604346392376_2174
```

```
20/11/15 08:40:06 WARN security.UserGroupInformation: PriviledgedActionException
```

```

as:lh2514 (auth:SIMPLE) cause:org.apache.hadoop.mapreduce.lib.input.InvalidInput
tException: Input path does not exist: hdfs://dumbo/user/lh2514/NYPD_Complaint_D
ata_Historic.csv
Exception in thread "main" org.apache.hadoop.mapreduce.lib.input.InvalidInputExc
eption: Input path does not exist: hdfs://dumbo/user/lh2514/NYPD_Complaint_Data_
Historic.csv
    at org.apache.hadoop.mapreduce.lib.input.FileInputFormat.singleThreadedL
istStatus(FileInputFormat.java:323)
    at org.apache.hadoop.mapreduce.lib.input.FileInputFormat.listStatus(File
InputFormat.java:265)
    at org.apache.hadoop.mapreduce.lib.input.FileInputFormat.get_splits(FileI
nputFormat.java:387)
    at org.apache.hadoop.mapreduce.JobSubmitter.writeNew_splits(JobSubmitter.
java:305)
    at org.apache.hadoop.mapreduce.JobSubmitter.write_splits(JobSubmitter.jav
a:322)
    at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitt
er.java:200)
    at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1307)
    at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1304)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
    at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInforma
tion.java:1924)
    at org.apache.hadoop.mapreduce.Job.submit(Job.java:1304)
    at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1325)
    at SelectFieldsRunner.main(SelectFieldsRunner.java:45)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.
java:62)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAcces
sorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:498)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:226)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:141)
[lh2514@login-2-1 ~]$ vi SelectFieldsRunner.java
[lh2514@login-2-1 ~]$ javac -classpath "$(yarn classpath)":.. -d . SelectFieldsRu
nner.java
Note: SelectFieldsRunner.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
[lh2514@login-2-1 ~]$ jar -cvf SelectFields.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: 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 SelectFields.jar SelectFieldsRunner /user/lh2514/class1/NYPD_Complaint_Data_Historic.csv /user/lh2514/class1/NYC_output1
20/11/15 08:42:48 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
20/11/15 08:42:49 INFO input.FileInputFormat: Total input paths to process : 1
20/11/15 08:42:49 INFO mapreduce.JobSubmitter: number of splits:2
20/11/15 08:42:49 INFO Configuration.deprecation: mapred.textoutputformat.separator is deprecated. Instead, use mapreduce.output.textoutputformat.separator
20/11/15 08:42:49 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1604346392376_2175
20/11/15 08:42:50 INFO impl.YarnClientImpl: Submitted application application_1604346392376_2175
20/11/15 08:42:50 INFO mapreduce.Job: The url to track the job: http://babar.es.its.nyu.edu:8088/proxy/application_1604346392376_2175/
20/11/15 08:42:50 INFO mapreduce.Job: Running job: job_1604346392376_2175
20/11/15 08:42:55 INFO mapreduce.Job: Job job_1604346392376_2175 running in uber mode : false
20/11/15 08:42:55 INFO mapreduce.Job:  map 0% reduce 0%
20/11/15 08:43:01 INFO mapreduce.Job:  map 100% reduce 0%
20/11/15 08:43:08 INFO mapreduce.Job:  map 100% reduce 100%
20/11/15 08:43:09 INFO mapreduce.Job: Job job_1604346392376_2175 completed successfully
20/11/15 08:43:09 INFO mapreduce.Job: Counters: 49
    File System Counters
        FILE: Number of bytes read=10213864
        FILE: Number of bytes written=20929825
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=265793063
        HDFS: Number of bytes written=52955815
        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)=33412
        Total time spent by all reduces in occupied slots (ms)=26028
        Total time spent by all map tasks (ms)=8353
        Total time spent by all reduce tasks (ms)=4338
        Total vcore-milliseconds taken by all map tasks=8353
        Total vcore-milliseconds taken by all reduce tasks=4338
        Total megabyte-milliseconds taken by all map tasks=34213888
        Total megabyte-milliseconds taken by all reduce tasks=26652672
    Map-Reduce Framework
        Map input records=1048576
```

Map output records=1048576
Map output bytes=54004391
Map output materialized bytes=10251739
Input split bytes=258
Combine input records=0
Combine output records=0
Reduce input groups=834706
Reduce shuffle bytes=10251739
Reduce input records=1048576
Reduce output records=1048576
Spilled Records=2097152
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=498
CPU time spent (ms)=19930
Physical memory (bytes) snapshot=2951094272
Virtual memory (bytes) snapshot=11212578816
Total committed heap usage (bytes)=4353687552

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=265792805

File Output Format Counters

Bytes Written=52955815

[lh2514@login-2-1 ~]\$ hdfs dfs -ls /user/lh2514/class1/NYC_output1

Found 2 items

-rw-r--r--+ 3 lh2514 users 0 2020-11-15 08:43 /user/lh2514/class1/NYC_output1/_SUCCESS
-rw-r--r--+ 3 lh2514 users 52955815 2020-11-15 08:43 /user/lh2514/class1/NYC_output1/part-r-00000

[lh2514@login-2-1 ~]\$ hdfs dfs -cat /user/lh2514/class1/NYC_output1/part-r-00000

,0:01:00,GRAND LARCENY,N.Y. POLICE DEPT
,0:15:00,PETIT LARCENY,N.Y. POLICE DEPT
,10:00:00,BURGLARY,N.Y. POLICE DEPT
,11:00:00,GRAND LARCENY,N.Y. POLICE DEPT
,11:00:00,HARRASSMENT 2,COMPLETED
,11:30:00,PETIT LARCENY,MISDEMEANOR
,11:30:00,PETIT LARCENY,N.Y. POLICE DEPT
,12:00:00,CRIMINAL MISCHIEF & RELATED OF,N.Y. POLICE DEPT
,12:00:00,GRAND LARCENY,N.Y. POLICE DEPT
,12:00:00,HARRASSMENT 2,VIOLATION
,12:00:00,OFF. AGNST PUB ORD SENSBLTY &,QUEENS
,13:00:00,CRIMINAL MISCHIEF & RELATED OF,FELONY
,13:15:00,SEX CRIMES,N.Y. POLICE DEPT
,13:23:00,OFF. AGNST PUB ORD SENSBLTY &,BRONX
,13:30:00,HARRASSMENT 2,COMPLETED

,13:40:00,HARRASSMENT 2,VIOLATION
,14:00:00,GRAND LARCENY,N.Y. POLICE DEPT
,14:00:00,GRAND LARCENY,N.Y. POLICE DEPT
,15:00:00,BURGLARY,N.Y. POLICE DEPT
,15:00:00,GRAND LARCENY,N.Y. POLICE DEPT
,15:00:00,PETIT LARCENY,MISDEMEANOR
,15:00:00,PETIT LARCENY,MISDEMEANOR
,15:00:00,PETIT LARCENY,N.Y. POLICE DEPT
,15:30:00,OFF. AGNST PUB ORD SENSBLTY &,BRONX
,16:00:00,OFF. AGNST PUB ORD SENSBLTY &,BROOKLYN
,16:00:00,PETIT LARCENY,N.Y. POLICE DEPT
,16:00:00,PETIT LARCENY,N.Y. POLICE DEPT
,16:16:00,MISCELLANEOUS PENAL LAW,MANHATTAN
,16:30:00,GRAND LARCENY,N.Y. POLICE DEPT
,16:30:00,OFF. AGNST PUB ORD SENSBLTY &,BRONX
,17:00:00,ASSAULT 3 & RELATED OFFENSES,BRONX
,17:00:00,PETIT LARCENY,MISDEMEANOR
,17:00:00,ROBBERY,N.Y. POLICE DEPT
,17:50:00,PETIT LARCENY,N.Y. POLICE DEPT
,18:00:00,HARRASSMENT 2,COMPLETED
,18:00:00,PETIT LARCENY,N.Y. POLICE DEPT
,18:30:00,PETIT LARCENY,MISDEMEANOR
,18:50:00,ASSAULT 3 & RELATED OFFENSES,N.Y. POLICE DEPT
,19:00:00,PETIT LARCENY,N.Y. POLICE DEPT
,19:10:00,HARRASSMENT 2,COMPLETED
,19:30:00,OFF. AGNST PUB ORD SENSBLTY &,QUEENS
,1:00:00,CRIMINAL MISCHIEF & RELATED OF,MISDEMEANOR
,1:45:00,FELONY ASSAULT,FELONY
,20:00:00,HARRASSMENT 2,COMPLETED
,21:00:00,CRIMINAL MISCHIEF & RELATED OF,N.Y. POLICE DEPT
,21:00:00,PETIT LARCENY,N.Y. POLICE DEPT
,21:13:00,HARRASSMENT 2,COMPLETED
,22:00:00,CRIMINAL MISCHIEF & RELATED OF,N.Y. POLICE DEPT
,22:00:00,CRIMINAL MISCHIEF & RELATED OF,N.Y. POLICE DEPT
,22:15:00,ASSAULT 3 & RELATED OFFENSES,BROOKLYN
,22:55:00,ASSAULT 3 & RELATED OFFENSES,MANHATTAN
,23:00:00,GRAND LARCENY,N.Y. POLICE DEPT
,23:55:00,ASSAULT 3 & RELATED OFFENSES,MANHATTAN
,2:00:00,ROBBERY,N.Y. POLICE DEPT
,2:40:00,PETIT LARCENY,N.Y. POLICE DEPT
,3:00:00,ROBBERY,N.Y. POLICE DEPT
,3:30:00,BURGLARY,FELONY
,4:20:00,PETIT LARCENY,MISDEMEANOR
,6:00:00,RAPE,N.Y. POLICE DEPT
,7:10:00,HARRASSMENT 2,COMPLETED
,8:15:00,VEHICLE AND TRAFFIC LAWS,MANHATTAN
,8:45:00,HARRASSMENT 2,VIOLATION
,9:00:00,GRAND LARCENY,N.Y. POLICE DEPT
,9:00:00,PETIT LARCENY,N.Y. POLICE DEPT
,9:15:00,FELONY ASSAULT,FELONY
09/16/1015,12:00:00,SEX CRIMES,N.Y. POLICE DEPT
09/26/1015,12:11:00,GRAND LARCENY,FELONY

1/1/1971,0:01:00,SEX CRIMES,N.Y. POLICE DEPT
1/1/1975,0:01:00,RAPE,QUEENS
1/1/1977,0:01:00,SEX CRIMES,MANHATTAN
1/1/1978,12:00:00,THEFT-FRAUD,N.Y. POLICE DEPT
1/1/1988,12:00:00,FRAUDS,N.Y. POLICE DEPT
1/1/1989,0:01:00,SEX CRIMES,N.Y. POLICE DEPT
1/1/1990,0:01:00,RAPE,QUEENS
1/1/1990,0:01:00,SEX CRIMES,MANHATTAN
1/1/1991,0:01:00,SEX CRIMES,BROOKLYN
1/1/1992,0:00:00,THEFT-FRAUD,N.Y. POLICE DEPT
1/1/1992,0:01:00,SEX CRIMES,MANHATTAN
1/1/1993,0:00:00,SEX CRIMES,N.Y. POLICE DEPT
1/1/1993,0:01:00,RAPE,BROOKLYN
1/1/1993,0:01:00,RAPE,QUEENS
1/1/1994,0:00:00,GRAND LARCENY,N.Y. HOUSING POLICE
1/1/1994,0:01:00,THEFT-FRAUD,N.Y. HOUSING POLICE
1/1/1995,0:01:00,OFF. AGNST PUB ORD SENSBLTY &,MANHATTAN
1/1/1995,12:00:00,MURDER & NON-NEGL. MANSLAUGHTER,BRONX
1/1/1996,0:01:00,RAPE,BROOKLYN
1/1/1997,12:00:00,FRAUDS,N.Y. HOUSING POLICE
1/1/1997,12:00:00,GRAND LARCENY,N.Y. POLICE DEPT
1/1/1997,1:00:00,SEX CRIMES,QUEENS
1/1/1997,22:00:00,RAPE,BRONX
1/1/1998,9:00:00,SEX CRIMES,N.Y. POLICE DEPT
1/1/1999,11:00:00,THEFT-FRAUD,N.Y. POLICE DEPT
1/1/1999,12:00:00,SEX CRIMES,N.Y. POLICE DEPT
1/1/1999,15:00:00,RAPE,BROOKLYN
1/1/1999,8:00:00,SEX CRIMES,STATEN ISLAND
1/1/1999,9:00:00,SEX CRIMES,QUEENS
1/1/2000,0:01:00,OTHER STATE LAWS (NON PENAL LA,N.Y. POLICE DEPT
1/1/2000,0:01:00,SEX CRIMES,N.Y. POLICE DEPT
1/1/2000,0:01:00,SEX CRIMES,QUEENS
1/1/2000,13:00:00,THEFT-FRAUD,N.Y. POLICE DEPT
1/1/2001,0:00:00,RAPE,BROOKLYN
1/1/2001,0:01:00,FRAUDS,N.Y. POLICE DEPT
1/1/2001,0:01:00,GRAND LARCENY,N.Y. POLICE DEPT
1/1/2001,0:01:00,GRAND LARCENY,N.Y. POLICE DEPT
1/1/2001,0:01:00,SEX CRIMES,BRONX
1/1/2001,0:01:00,SEX CRIMES,BRONX
1/1/2001,0:01:00,SEX CRIMES,BROOKLYN
1/1/2001,0:01:00,SEX CRIMES,BROOKLYN
1/1/2001,0:01:00,SEX CRIMES,MANHATTAN
1/1/2001,12:00:00,GRAND LARCENY,N.Y. HOUSING POLICE
1/1/2001,15:00:00,GRAND LARCENY,N.Y. POLICE DEPT
1/1/2001,23:00:00,SEX CRIMES,N.Y. POLICE DEPT
1/1/2001,7:00:00,FRAUDS,N.Y. HOUSING POLICE
1/1/2001,9:00:00,SEX CRIMES,MANHATTAN
1/1/2002,0:01:00,GRAND LARCENY,N.Y. POLICE DEPT
1/1/2002,0:01:00,RAPE,BROOKLYN
1/1/2002,0:01:00,SEX CRIMES,N.Y. POLICE DEPT
1/1/2002,0:01:00,SEX CRIMES,N.Y. POLICE DEPT
1/1/2002,0:01:00,THEFT-FRAUD,N.Y. HOUSING POLICE