

# CPR E 419 - Lab 4

## 1. pig -x -local exp1.pig

```
Job Stats (time in seconds):
JobId  Maps  Reduces  MaxMapTime  MinMapTime  AvgMapTime  MedianMapTime  MaxReduceTime  MinReduceTime  AvgReduceTime  MedianReduceTime  Alias  Feature Outputs
job_local1113698577_0004  1  1  n/a  n/a  n/a  n/a  n/a  n/a  n/a  n/a  n/a  ordered_aland  file:///home/cpre419/Downloads/lab4_exp1_output,
job_local157904908_0003  1  1  n/a  n/a  n/a  n/a  n/a  n/a  n/a  n/a  n/a  ordered_aland  ORDER_BY_COMBINER
job_local1666339335_0002  1  1  n/a  n/a  n/a  n/a  n/a  n/a  n/a  n/a  n/a  ordered_aland  SAMPLER
job_local71261153_0001  1  1  n/a  n/a  n/a  n/a  n/a  n/a  n/a  n/a  n/a  input_lines,states,total  GROUP_BY_COMBINER

Input(s):
Successfully read 148086 records from: "file:///home/cpre419/Downloads/gaz_tracts_national.txt"

Output(s):
Successfully stored 10 records in: "file:///home/cpre419/Downloads/lab4_exp1_output"

Counters:
Total records written : 10
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local71261153_0001 -> job_local1666339335_0002,
job_local1666339335_0002 -> job_local157904908_0003,
job_local157904908_0003 -> job_local1113698577_0004,
job_local1113698577_0004

2021-03-17 23:43:52,631 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-17 23:43:52,633 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-17 23:43:52,637 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-17 23:43:52,648 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-17 23:43:52,668 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-17 23:43:52,671 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-17 23:43:52,701 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-17 23:43:52,711 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-17 23:43:52,712 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-17 23:43:52,717 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-17 23:43:52,721 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-17 23:43:52,722 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-17 23:43:52,756 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 6 time(s).
2021-03-17 23:43:52,756 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - Encountered Warning ACCESSING_NON_EXISTENT_FIELD 666027 time(s).
2021-03-17 23:43:52,756 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - Encountered Warning TOO_LARGE_FOR_INT 55 time(s).
2021-03-17 23:43:52,840 [main] INFO org.apache.pig.Main - Pig script completed in 15 seconds and 233 milliseconds (15233 ms)
```

AK	1477953211577
TX	676586997978
CA	403466310059
MT	376961878670
NM	314160748240
AZ	294207314414
NV	284331937541
CO	268431246426
WY	251470069067
OR	248607802255

## 2. pig -x -mapreduce exp2.pig

```
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Output Format Counters
  Bytes Written=0
2021-03-19 06:18:42,661 [pool-21-thread-1] INFO org.apache.hadoop.mapred.LocalJobRunner - Finishing task: attempt_local1992644686_0005_r_000000_0
2021-03-19 06:18:42,661 [Thread-178] INFO org.apache.hadoop.mapred.LocalJobRunner - reduce task executor complete.
2021-03-19 06:18:42,900 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_local1992644686_0005]
2021-03-19 06:18:47,404 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-19 06:18:47,406 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-19 06:18:47,408 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-19 06:18:47,428 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2021-03-19 06:18:47,466 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:

HadoopVersion  PigVersion  UserId  StartedAt      FinishedAt      Features
3.1.3    0.17.0    cpre419  2021-03-19 06:18:10    2021-03-19 06:18:47    GROUP_BY,ORDER_BY,DISTINCT,FILTER,LIMIT

Success!

Job Stats (time in seconds):
JobId  Maps  Reduces  MaxMapTime  MinMapTime  AvgMapTime  MedianMapTime  MaxReduceTime  MinReduceTime  AvgReduceTime  MedianReduceTime  Alias  Feature Outputs
job_local1729768507_0002  1  1  n/a  n/a  n/a  n/a  n/a  n/a  groupedlps, total  GROUP_BY,COMBINER
job_local1729768507_0002  1  1  n/a  n/a  n/a  n/a  n/a  n/a  sortedlps  ORDER_BY,COMBINER
job_local1992644686_0005  1  1  n/a  n/a  n/a  n/a  n/a  n/a  sortedlps  /Lab4/exp2/output,
job_local1838425969_0001  1  1  n/a  n/a  n/a  n/a  n/a  n/a  filteredlines, inputlines, lppairs  DISTINCT
job_local1906885772_0003  1  1  n/a  n/a  n/a  n/a  n/a  n/a  sortedlps  SAMPLER

Input(s):
Successfully read 600000 records (95202500 bytes) from: "/data/network_trace"

Output(s):
Successfully stored 10 records (57645411 bytes) in: "/lab4/exp2/output"

Counters:
Total records written : 10
Total bytes written : 57645411
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1838425969_0001 -> job_local1729768507_0002,
job_local1729768507_0002 -> job_local1906885772_0003,
job_local1906885772_0003 -> job_local174889311_0004,
job_local174889311_0004 -> job_local1992644686_0005,
job_local1992644686_0005
```

```
10.31.24.241    273
192.168.1.4     114
10.31.4.211     70
10.31.4.188     52
107.152.26.197  4
17.134.127.250  4
162.125.34.137  4
162.125.33.7    4
162.125.18.133  4
107.152.25.197  4
```

### 3. pig -x -mapreduce exp3.pig

```
job_local949293759_0003 1 1 n/a n/a n/a n/a n/a n/a n/a ordered SAMPLER

Input(s):
Successfully read 600000 records from: "/data/ip_trace"
Successfully read 600000 records from: "/data/raw_block"

Output(s):
Successfully stored 419801 records (23794367 bytes) in: "/lab4/exp3/firewall"
Successfully stored 1613 records (149776181 bytes) in: "/lab4/exp3/output"

Counters:
Total records written : 421414
Total bytes written : 173570548
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1294713995_0001 -> job_local500039020_0002,
job_local500039020_0002 -> job_local949293759_0003,
job_local949293759_0003 -> job_local2090719570_0004,
job_local2090719570_0004

2021-03-19 11:37:30,159 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-19 11:37:30,171 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-19 11:37:30,191 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-19 11:37:30,198 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-19 11:37:30,236 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-19 11:37:30,239 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-19 11:37:30,282 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-19 11:37:30,289 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-19 11:37:30,298 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-19 11:37:30,329 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-19 11:37:30,331 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-19 11:37:30,339 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2021-03-19 11:37:30,372 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreducelayer.MapReduceLauncher - Success!
2021-03-19 11:37:30,543 [main] INFO org.apache.pig.Main - Pig script completed in 1 minute, 18 seconds and 609 milliseconds (78609 ms)
cpre419@cpre419-VirtualBox:~/Downloads$ hadoop fs -copyToLocal /lab4/exp3/output .
copyToLocal: 'output/' SUCCESS: File exists
copyToLocal: 'output/part-r-00000': File exists
cpre419@cpre419-VirtualBox:~/Downloads$ hadoop fs -copyToLocal /lab4/exp3/output .
2021-03-19 11:38:34,477 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
cpre419@cpre419-VirtualBox:~/Downloads$ hadoop fs -copyToLocal /lab4/exp3/firewall .
2021-03-19 11:38:46,873 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
cpre419@cpre419-VirtualBox:~/Downloads$
```

In experiment 3, I have included part 3a and 3b tasks under the overall exp3.pig file, as was pointed out by a comment under Piazza.

```
15:28:46.539147 0 10.31.27.252 224.0.0.102 Blocked
15:28:47.479137 1 10.31.26.186 224.0.0.251 Blocked
15:28:47.479217 2 10.31.26.216 224.0.0.251 Blocked
15:28:48.105882 3 10.31.27.253 224.0.0.102 Blocked
15:28:48.628287 5 10.10.251 224.0.0 Blocked
15:28:49.672578 8 10.31.26.216 224.0.0.251 Blocked
15:28:50.194885 10 10.31.24.175 224.0.0.251 Blocked
15:28:50.299328 11 10.33.14.68 224.0.0.251 Blocked
15:28:50.403757 12 10.33.14.68 224.0.0.251 Blocked
15:28:50.508240 13 10.31.26.216 224.0.0.251 Blocked
15:28:50.717141 14 10.31.27.253 224.0.0.102 Blocked
15:28:50.924455 17 129.186.140.200 10.31.24.241 Blocked
15:28:50.925952 19 10.33.14.37 224.0.0.251 Blocked
15:28:51.436964 20 10.31.24 224.0.0 Blocked
15:28:51.813457 21 10.31.24.241 172.217.1.69 Blocked
15:28:51.848832 25 172.217.1.69 10.31.24.241 Blocked
15:28:51.848851 27 172.217.1.69 10.31.24.241 Blocked
15:28:51.848856 28 172.217.1.69 10.31.24.241 Blocked
15:28:51.848861 29 172.217.1.69 10.31.24.241 Blocked
15:28:51.848904 30 10.31.24.241 172.217.1.69 Blocked
```

13.249.140.62	73223
10.31.24.241	64393
23.35.205.78	46495
10.31.4.211	11814
151.101.148.133	11473
192.168.1.4	11443
130.211.46.84	10590
10.31.4.188	8567
10.31.6.140	4959
10.33.15.125	4225
100.24.180.47	3864
10.31.27.253	3816
10.31.27.252	3764
162.125.18.133	2526
10.31.7.253	2381
10.31.7.252	2347
45.33.67.183	2133
10.31.4.32	1987
129.186.140.200	1809
129.186.78.200	1657
10.31.24.165	1644