Lab 1

Load weather data file and apply different transformations on it.

Data Definition:

Column 1: Weather Station

Column 2: Date(yyyyMMDD)

Column 3: Observation Type

Column 4: Temperature

TEMP_LOAD=LOAD 'weather.txt' USING PigStorage(',') AS (station:chararray,datestr:chararray,obstype:chararray,temp:int);

```
grunt> TEMP_LOAD = LOAD 'weather.txt' USING PigStorage(',') AS (station:chararray,datestr:chararray,obstype:chararray,temp:int);
(EZE00100082,20181225,TMAX,21)
(EZE00100082,20181225,TMIN,11)
(ITE00100554,20181226,TMAX,38)
(ITE00100554,20181226,TMIN,25)
(GM000010962,20181226,PRCP,0)
(EZE00100082,20181226,TMAX,36)
(EZE00100082,20181226,TMIN,8)
(ITE00100554,20181227,TMAX,10)
(ITE00100554,20181227,TMIN,9)
(GM000010962,20181227,PRCP,0)
(EZE00100082,20181227,TMAX,47)
(EZE00100082,20181227,TMIN,20)
(ITE00100554,20181228,TMAX,35)
(ITE00100554,20181228,TMIN,25)
(GM000010962,20181228,PRCP,7)
(EZE00100082,20181228,TMAX,43)
(EZE00100082,20181228,TMIN,7)
(ITE00100554,20181229,TMAX,50)
(ITE00100554,20181229,TMIN,16)
(GM000010962,20181229,PRCP,9)
(EZE00100082,20181229,TMAX,45)
(EZE00100082,20181229,TMIN,-3)
(ITE00100554,20181230,TMAX,50)
(ITE00100554,20181230,TMIN,31)
(GM000010962,20181230,PRCP,13)
(EZE00100082,20181230,TMAX,96)
(EZE00100082,20181230,TMIN,14)
(ITE00100554,20181231,TMAX,50)
(ITE00100554,20181231,TMIN,25)
(GM000010962,20181231,PRCP,16)
(EZE00100082,20181231,TMAX,14)
(EZE00100082,20181231,TMIN,-4)
grunt>
```

TEMP_GROUP= GROUP TEMP_LOAD BY station;

DUMP TEMP_GROUP;

```
grunt> TEMP_GROUP = GROUP TEMP_LOAD by station;
```

```
grunt> dump TEMP_GROUP;
2019-12-08 10:11:44,220 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: GROUP_BY
2019-12-08 10:11:44,222 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColu
mnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForE
achFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}
2019-12-08 10:11:44,232 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2019-12-08 10:11:44,242 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2019-12-08 10:11:44,242 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2019-12-08 10:11:44,276 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2019-12-08 10:11:44,279 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2019-12-08 10:11:44,300 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not
 set, set to default 0.3
2019-12-08 10:11:44,300 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required r
educers.
2019-12-08 10:11:44,305 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using reducer estimator: org.apache.pig.backend.h
adoop.executionengine.mapReduceLayer.InputSizeReducerEstimator
2019-12-08 10:11:44,308 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator - BytesPerReducer=1000000000 maxReducers=999
 totalInputFileSize=53592
2019-12-08 10:11:44,308 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1
2019-12-08 10:11:44,694 [DataStreamer for file /tmp/temp-996373873/tmp491104500/jdo-api-3.0.1.jar] WARN  org.apache.hadoop.hdfs.DFSClient - Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1281)
        at java.lang.Thread.join(Thread.java:1355)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
```

TEMP_MIN=FOREACH TEMP_GROUP GENERATE TEMP_LOAD.station, MIN(TEMP_LOAD.temp);

DUMP TEMP_MIN;

```
grunt> TEMP_MIN = FOREACH TEMP_GROUP GENERATE TEMP_LOAD.station,MIN(TEMP_LOAD.temp)
>> ;
grunt>
```

```
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554
),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554),(ITE00100554)},-148)
```

TEMP_FILTER=FILTER TEMP_LOAD BY temp > 0;

ILLUSTRATE TEMP_FILTER;

```
grunt> TEMP_FILTER = FILTER TEMP_LOAD
>>
>>
>>
>>
>>
>>
>>
>>
>>
>> BY temp>0;
grunt>
```

```
grunt> dump TEMP_FILTER;
2019-12-08 10:18:15,876 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: FILTER
2019-12-08 10:18:15,878 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColu
mnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForE
achFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}
2019-12-08 10:18:15,896 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2019-12-08 10:18:15,900 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2019-12-08 10:18:15,900 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2019-12-08 10:18:15,968 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2019-12-08 10:18:15,974 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2019-12-08 10:18:16,005 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not
 set, set to default 0.3
2019-12-08 10:18:16,115 [DataStreamer for file /tmp/temp-996373873/tmp1452854280/libthrift-0.9.3.jar] WARN  org.apache.hadoop.hdfs.DFSClient - Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1281)
        at java.lang.Thread.join(Thread.java:1355)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
```

```
(EZE00100082,20181224,TMAX,30)
(EZE00100082,20181224,TMIN,6)
(ITE00100554,20181225,TMAX,25)
(ITE00100554,20181225,TMIN,13)
(EZE00100082,20181225,TMAX,21)
(EZE00100082,20181225,TMIN,11)
(ITE00100554,20181226,TMAX,38)
(ITE00100554,20181226,TMIN,25)
(EZE00100082,20181226,TMAX,36)
(EZE00100082,20181226,TMIN,8)
(ITE00100554,20181227,TMAX,10)
(ITE00100554,20181227,TMIN,9)
(EZE00100082,20181227,TMAX,47)
(EZE00100082,20181227,TMIN,20)
(ITE00100554,20181228,TMAX,35)
(ITE00100554,20181228,TMIN,25)
(GM000010962,20181228,PRCP,7)
(EZE00100082,20181228,TMAX,43)
(EZE00100082,20181228,TMIN,7)
(ITE00100554,20181229,TMAX,50)
(ITE00100554,20181229,TMIN,16)
(GM000010962,20181229,PRCP,9)
(EZE00100082,20181229,TMAX,45)
(ITE00100554,20181230,TMAX,50)
(ITE00100554,20181230,TMIN,31)
(GM000010962,20181230,PRCP,13)
(EZE00100082,20181230,TMAX,96)
(EZE00100082,20181230,TMIN,14)
(ITE00100554,20181231,TMAX,50)
(ITE00100554,20181231,TMIN,25)
(GM000010962,20181231,PRCP,16)
(EZE00100082,20181231,TMAX,14)
```

STORE TEMP_MIN INTO 'hdfs://localhost:9000/temp_output' USING PigStorage(',');

```
grunt> STORE TEMP_MIN INTO '/temp_output' USING PigStorage(',');
2019-12-08 10:21:06,623 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: GROUP_BY
2019-12-08 10:21:06,624 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColu
mnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForE
achFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}
2019-12-08 10:21:06,634 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2019-12-08 10:21:06,637 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2019-12-08 10:21:06,637 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2019-12-08 10:21:06,672 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2019-12-08 10:21:06,674 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2019-12-08 10:21:06,695 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not
 set, set to default 0.3
2019-12-08 10:21:06,695 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required r
educers.
2019-12-08 10:21:06,695 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using reducer estimator: org.apache.pig.backend.h
adoop.executionengine.mapReduceLayer.InputSizeReducerEstimator
2019-12-08 10:21:06,697 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator - BytesPerReducer=1000000000 maxReducers=999
 totalInputFileSize=53592
2019-12-08 10:21:06,697 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1
2019-12-08 10:21:07,021 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job7755396830500266271.jar
2019-12-08 10:21:11,226 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job7755396830500266271.jar created
2019-12-08 10:21:11,234 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2019-12-08 10:21:11,235 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2019-12-08 10:21:11,235 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2019-12-08 10:21:11,235 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []


HadoopVersion   PigVersion      UserId   StartedAt       FinishedAt      Features
2.6.0-cdh5.13.0 0.12.0-cdh5.13.0         cloudera        2019-12-08 10:21:06    2019-12-08 10:21:46     GROUP_BY

Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTIme      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime        Alias  F
eature  Outputs
job_1575799937674_0008  1       1       7       7       7       7       8       8       8       8       TEMP_GROUP,TEMP_LOAD,TEMP_MIN   GROUP_BY        /temp_output,

Input(s):
Successfully read 1825 records (53968 bytes) from: "hdfs://quickstart.cloudera:8020/user/cloudera/weather.txt"

Output(s):
Successfully stored 3 records (25568 bytes) in: "/temp_output"

Counters:
Total records written : 3
Total bytes written : 25568
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1575799937674_0008


2019-12-08 10:21:47,037 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt>
```

Lab 2.

Load sports data and apply transformations on it.

Data set used: PIG_SAMPLE_DATA

2.1 Load file into pig

```
grunt> Sdata = LOAD 'pig_sample_data.csv' USING PigStorage(',') AS (id:int,name:chararray,email:chararray,job:chararray,state:chararray);
grunt> dump Sdata;
(19649,Elijah Collis,Elijah_Collis789@iatim.tech,Healthcare Specialist,Michigan)
(19650,Abdul Robinson,Abdul_Robinson6234@nimogy.biz,Banker,Ohio)
(19651,Josh Uttridge,Josh_Uttridge5681@jiman.org,Business Broker,Alaska)
(19652,Chuck Vince,Chuck_Vince44@ubusive.com,Production Painter,Arkansas)
(19653,Hayden Bolton,Hayden_Bolton9345@bretoux.com,Laboratory Technician,Nevada)
(19654,Julia Ingham,Julia_Ingham145@famism.biz,Baker,Montana)
(19655,Camden Spencer,Camden_Spencer7134@qater.org,Chef Manager,Utah)
(19656,Ronald Isaac,Ronald_Isaac7071@womeona.net,Lecturer,Iowa)
(19657,Noah Lloyd,Noah_Lloyd2973@nickia.com,IT Support Staff,Vermont)
(19658,Leroy Shields,Leroy_Shields3756@typill.biz,Dentist,Ohio)
(19659,Carissa Hopkins,Carissa_Hopkins2378@brety.org,Investment  Advisor,South Dakota)
(19660,Leroy Nanton,Leroy_Nanton4947@tonsy.org,HR Coordinator,New Hampshire)
(19661,Nancy Plant,Nancy_Plant7865@womeona.net,Production Painter,Colorado)
(19662,Havana Oliver,Havana_Oliver3667@bulaffy.com,Designer,South Dakota)
(19663,Christine Morley,Christine_Morley248@acrit.org,Associate Professor,Kentucky)
(19664,Kurt Ripley,Kurt_Ripley7942@ubusive.com,Biologist,Indiana)
(19665,Mike Graham,Mike_Graham4867@gompie.com,Project Manager,Oregon)
(19666,Nick Stanton,Nick_Stanton2968@kideod.biz,Accountant,Louisiana)
(19667,Melanie Ballard,Melanie_Ballard8195@dionrab.com,Food Technologist,Alaska)
(19668,Carolyn Ripley,Carolyn_Ripley8373@muall.tech,Software Engineer,Ohio)
(19669,Joseph Nielson,Joseph_Nielson1826@extex.org,Retail Trainee,New Mexico)
(19670,Daniel Lane,Daniel_Lane3522@atink.com,Call Center Representative,Tennessee)
(19671,Angelina Garcia,Angelina_Garcia8958@bungar.biz,Banker,Kansas)
(19672,Andrea Mackenzie,Andrea_Mackenzie2617@iatim.tech,Lecturer,Louisiana)
(19673,Ada Wellington,Ada_Wellington9182@supunk.biz,Dentist,Minnesota)
(19674,Harry Blythe,Harry_Blythe4638@sveldo.biz,Accountant,Oregon)
(19675,Kieth Stuart,Kieth_Stuart9343@gmail.com,Doctor,Hawaii)
(19676,Matthew Dobson,Matthew_Dobson7219@deavo.com,Bookkeeper,Utah)
(19677,Payton Foxley,Payton_Foxley3638@iatim.tech,Software Engineer,Mississippi)
(19678,Alan Poole,Alan_Poole7571@gompie.com,Staffing Consultant,Washington)
(19679,Rose Kaur,Rose_Kaur8172@vetan.org,Fabricator,Connecticut)
(19680,Rufus Middleton,Rufus_Middleton2656@gmail.com,Paramedic,Tennessee)
(19681,Hanna Norman,Hanna_Norman6787@twipet.com,Loan Officer,Michigan)
```

```
cloudera@quickstart:~
```

2.2 find total medals won per year

Updated question: Find total no ppl where job title = accountant

```
grunt> accountant = FILTER Sdata by job == 'Accountant';
grunt>
```

```
(18209,Liam Drew,Liam_Drew8844@cispeto.com,Accountant,Indiana)
(18239,Alan Malone,Alan_Malone2032@elnee.tech,Accountant,Arizona)
(18266,Chester Price,Chester_Price7705@irrepsy.com,Accountant,Tennessee)
(18269,Robyn Douglas,Robyn_Douglas3190@brety.org,Accountant,Virginia)
(18350,Aurelia Hudson,Aurelia_Hudson7961@ovock.tech,Accountant,Arkansas)
(18364,Johnny Olson,Johnny_Olson4649@bauros.biz,Accountant,Oklahoma)
(18406,Charlotte Cavanagh,Charlotte_Cavanagh4982@nickia.com,Accountant,Pennsylvania)
(18435,Cherish Collins,Cherish_Collins5561@kideod.biz,Accountant,South Carolina)
(18466,Enoch Newman,Enoch_Newman4272@corti.com,Accountant,Arizona)
(18521,Maggie Notman,Maggie_Notman611@famism.biz,Accountant,Ohio)
(18569,Zara Edler,Zara_Edler8332@bulaffy.com,Accountant,Idaho)
(18641,Chris Rogan,Chris_Rogan8839@twipet.com,Accountant,Delaware)
(18652,Harry Farmer,Harry_Farmer4565@ubusive.com,Accountant,Colorado)
(18690,Chloe Clark,Chloe_Clark6401@bauros.biz,Accountant,New Mexico)
(18713,Enoch Robertson,Enoch_Robertson4497@nickia.com,Accountant,Nevada)
(18758,Chelsea Jefferson,Chelsea_Jefferson7993@elnee.tech,Accountant,Connecticut)
(18887,Elijah Bell,Elijah_Bell9633@infotech44.tech,Accountant,South Dakota)
(18897,Hazel Lowe,Hazel_Lowe4991@dionrab.com,Accountant,Alaska)
(18914,Gwenyth Stewart,Gwenyth_Stewart7750@typill.biz,Accountant,Ohio)
(18926,Janelle Ring,Janelle_Ring8574@extex.org,Accountant,Mississippi)
(18934,Chester Tyrrell,Chester_Tyrrell830@acrit.org,Accountant,Colorado)
(18963,Elena Mullins,Elena_Mullins1575@ubusive.com,Accountant,Washington)
(19237,Willow Purvis,Willow_Purvis4646@deavo.com,Accountant,Vermont)
(19397,Alma Jefferson,Alma_Jefferson3831@guentu.biz,Accountant,New Jersey)
(19544,Brooklyn Logan,Brooklyn_Logan1979@zorer.org,Accountant,Alabama)
(19646,Rhea Vangness,Rhea_Vangness5774@muall.tech,Accountant,Hawaii)
(19666,Nick Stanton,Nick_Stanton2968@kideod.biz,Accountant,Louisiana)
(19674,Harry Blythe,Harry_Blythe4638@sveldo.biz,Accountant,Oregon)
(19740,Vanessa Ulyatt,Vanessa_Ulyatt1627@twipet.com,Accountant,West Virginia)
(19911,Peter Wright,Peter_Wright6467@famism.biz,Accountant,Kentucky)
(19985,Ron Neal,Ron_Neal8605@fuliss.net,Accountant,North Dakota)
(19996,Jacqueline Kidd,Jacqueline_Kidd4686@gmail.com,Accountant,Pennsylvania)
grunt>
```

cloudera@quickstart:~

2.3 find total medals won by each athlete

Updated question: Total no of employeers from state = ohio

```
grunt> ohio = FILTER Sdata by state == 'Ohio';
grunt> dump ohio;
```

```
(18693,Enoch Grady,Enoch_Grady317@joiniaa.com,Bookkeeper,Ohio)
(18700,Bryon Bryant,Bryon_Bryant8554@liret.org,Electrician,Ohio)
(18720,Aeris Selby,Aeris_Selby4546@joiniaa.com,Budget Analyst,Ohio)
(18737,Stephanie Yarwood,Stephanie_Yarwood7623@grannar.com,Fabricator,Ohio)
(18780,Russel Ward,Russel_Ward3172@vetan.org,Restaurant Manager,Ohio)
(18883,Rosalee Whinter,Rosalee_Whinter4031@nickia.com,Cash Manager,Ohio)
(18914,Gwenyth Stewart,Gwenyth_Stewart7750@typill.biz,Accountant,Ohio)
(18920,Mike Walton,Mike_Walton5745@extex.org,Auditor,Ohio)
(18947,Charlize Evans,Charlize_Evans3929@iatim.tech,Executive Director,Ohio)
(18969,Diane Garcia,Diane_Garcia7941@gmail.com,Clerk,Ohio)
(19100,Elly Slater,Elly_Slater2692@sheye.org,Baker,Ohio)
(19246,Hayden Wild,Hayden_Wild4084@joiniaa.com,Production Painter,Ohio)
(19277,Angelique Gibson,Angelique_Gibson2779@yahoo.com,Biologist,Ohio)
(19293,Sage Atkinson,Sage_Atkinson481@iatim.tech,Business Broker,Ohio)
(19413,Kate Shepherd,Kate_Shepherd5137@kideod.biz,Bellman,Ohio)
(19433,Aiden Stone ,Aiden_Stone 6971@kideod.biz,Machine Operator,Ohio)
(19452,Nick Foxley,Nick_Foxley3868@nimogy.biz,Electrician,Ohio)
(19462,Phillip Niles,Phillip_Niles6020@tonsy.org,Service Supervisor,Ohio)
(19604,Clarissa Foxley,Clarissa_Foxley5102@jiman.org,Biologist,Ohio)
(19650,Abdul Robinson,Abdul_Robinson6234@nimogy.biz,Banker,Ohio)
(19658,Leroy Shields,Leroy_Shields3756@typill.biz,Dentist,Ohio)
(19668,Carolyn Ripley,Carolyn_Ripley8373@muall.tech,Software Engineer,Ohio)
(19742,Harvey Smith,Harvey_Smith2780@eirey.tech,Chef Manager,Ohio)
(19752,Rick Hopkinson,Rick_Hopkinson697@elnee.tech,Clerk,Ohio)
(19755,Nicholas Moore,Nicholas_Moore4308@mafthy.com,Machine Operator,Ohio)
(19769,Ryan Cox,Ryan_Cox8988@atink.com,Banker,Ohio)
(19773,Johnny Oliver,Johnny_Oliver7261@irrepsy.com,Cook,Ohio)
(19882,Russel Robe,Russel_Robe2846@elnee.tech,Global Logistics Supervisor,Ohio)
(19918,Domenic Gavin,Domenic_Gavin4248@bretoux.com,Systems Administrator,Ohio)
(19950,Russel Mccormick,Russel_Mccormick6982@acrit.org,Auditor,Ohio)
(19955,Leroy Burnley,Leroy_Burnley2514@kideod.biz,Call Center Representative,Ohio)
(19970,Kassandra Heaton,Kassandra_Heaton2233@deons.tech,Ambulatory Nurse,Ohio)
grunt> █
```

2.4 Save 2.3 results in HDFS.

```
grunt> STORE ohio  INTO '/ohio' USING PigStorage(',');
2019-12-08 11:21:20,263 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2019-12-08 11:21:20,263 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2019-12-08 11:21:20,449 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: FILTER
2019-12-08 11:21:20,450 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColu
mnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForE
achFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}
2019-12-08 11:21:20,461 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2019-12-08 11:21:20,464 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2019-12-08 11:21:20,464 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1

HadoopVersion  PigVersion     UserId  StartedAt       FinishedAt      Features
2.6.0-cdh5.13.0 0.12.0-cdh5.13.0       cloudera        2019-12-08 11:21:20    2019-12-08 11:21:56    FILTER

Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTIme      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime        Alias  F
eature  Outputs
job_1575799937674_0018  1       0       10      10      10      10      n/a     n/a     n/a     n/a     Sdata,ohio      MAP_ONLY        /ohio,

Input(s):
Successfully read 20001 records (1445792 bytes) from: "hdfs://quickstart.cloudera:8020/user/cloudera/pig_sample_data.csv"

Output(s):
Successfully stored 419 records (28112 bytes) in: "/ohio"

Counters:
Total records written : 419
Total bytes written : 28112
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1575799937674_0018


2019-12-08 11:21:56,842 [main] WARN  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSIO
N_FAILED 1 time(s).
2019-12-08 11:21:56,842 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt>
```