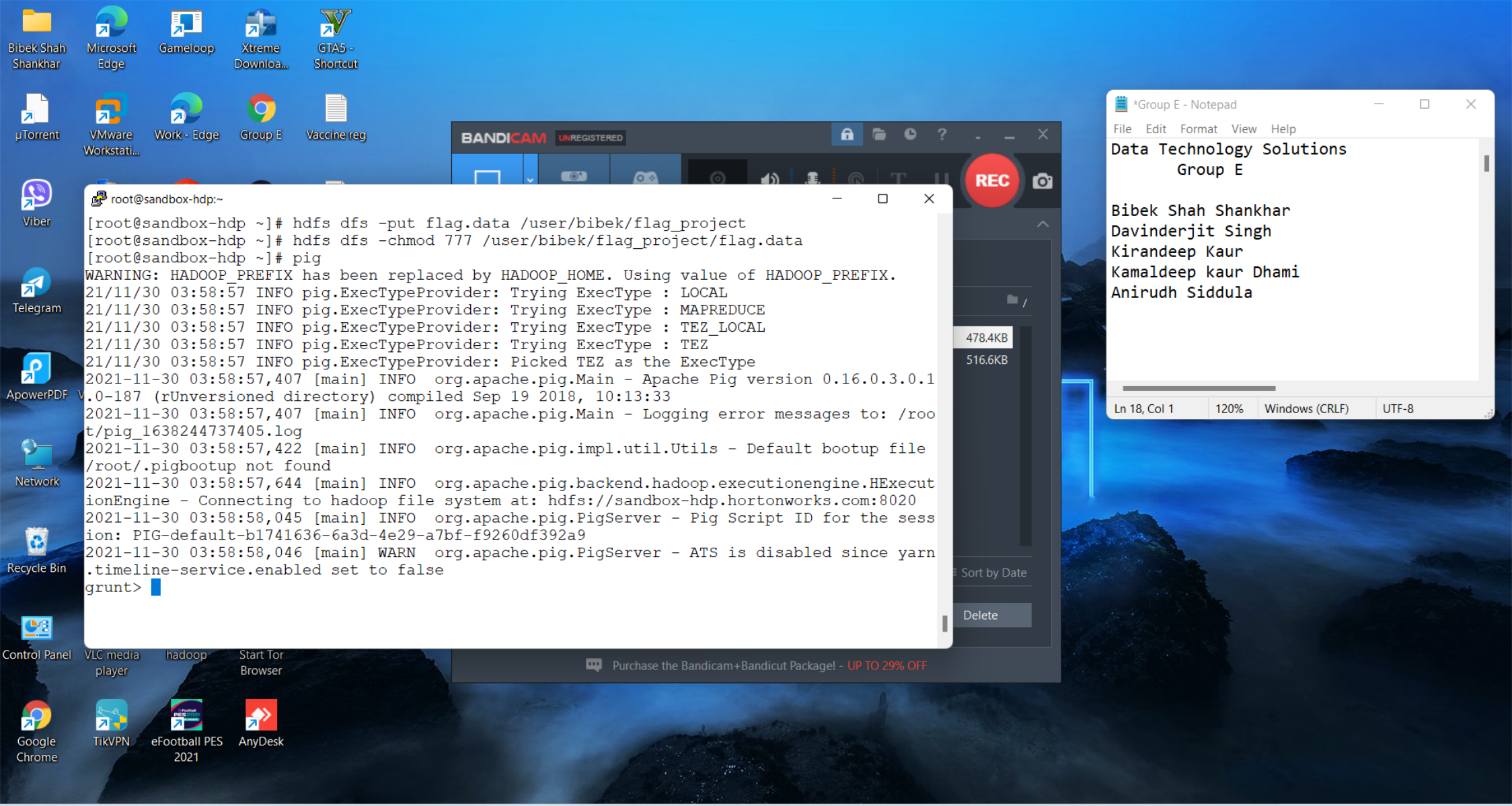
root@sandbox-hdp:~
[root@sandbox-hdp ~]# hdfs dfs -put flag.data /user/bibek/flag_project
[root@sandbox-hdp ~]# hdfs dfs -chmod 777 /user/bibek/flag_project/flag.data
[root@sandbox-hdp ~]# pig
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
21/11/30 03:58:57 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
21/11/30 03:58:57 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
21/11/30 03:58:57 INFO pig.ExecTypeProvider: Trying ExecType : TEZ_LOCAL
21/11/30 03:58:57 INFO pig.ExecTypeProvider: Trying ExecType : TEZ
21/11/30 03:58:57 INFO pig.ExecTypeProvider: Picked TEZ as the ExecType
2021-11-30 03:58:57,407 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0.3.0.1
.0-187 (rUnversioned directory) compiled Sep 19 2018, 10:13:33
2021-11-30 03:58:57,407 [main] INFO org.apache.pig.Main - Logging error messages to: /roo
t/pig_1638244737405.log
2021-11-30 03:58:57,422 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file
/root/.pigbootup not found
2021-11-30 03:58:57,644 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecut
ionEngine - Connecting to hadoop file system at: hdfs://sandbox-hdp.hortonworks.com:8020
2021-11-30 03:58:58,045 [main] INFO org.apache.pig.PigServer - Pig Script ID for the sess
ion: PIG-default-b1741636-6a3d-4e29-a7bf-f9260df392a9
2021-11-30 03:58:58,046 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn
.timeline-service.enabled set to false
grunt>
*Group E - Notepad
File Edit Format View Help
Data Technology Solutions
Group E
Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula
Ln 18, Col 1
120%
Windows (CRLF)
UTF-8
BANDICAM UNREGISTERED
REC
478.4KB
516.6KB
Sort by Date
Delete
Purchase the Bandicam+Bandicut Package! - UP TO 29% OFF

root@sandbox-hdp:~

```
grunt> flag_data = LOAD '/user/bibek/flag_project/flag.data' USING PigStorage(',') as (nam
e:chararray, landmass:int, zone:int, area:chararray, population:chararray, language:int, r
eligion:int, bars:int, stripes:int, colours:int, red:int, green:int, blue:int, gold:int, w
hite:int, black:int, orange:int, mainhue:chararray, circles:int, crosses:int, saltires:int
, quarters:int, sunstars:int, crescent:int, triangle:int, icon:int, animate:int, text:int,
 topleft:chararray, botright:chararray);
grunt> grouped_data = GROUP flag_data by landmass;
grunt>
grunt> count_data = FOREACH grouped_data GENERATE group,COUNT(flag_data);
grunt> DUMP count_data;
```

*Group E - Notepad

File Edit Format View Help

```
Data Technology Solutions
       Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula
A. Count number of countries based on
   landmass
landmass :
1=N.America, 2=S.America, 3=Europe,
4=Africa, 4=Asia, 6=Oceania
```

Ln 7, Col 17 | 120% | Windows (CRLF) | UTF-8

8:41 AM
01-Dec-21

root@sandbox-hdp:~

```
tesRead FileBytesWritten  HdfsBytesRead HdfsBytesWritten Alias  Feature Outputs
scope-103          1          1            194                  0            194
     64              109          15240                0 count_data,flag_data,grouped_data
scope-104          1          1              0                  6              6
    109                0             0               47 count_data     GROUP_BY        hd
fs://sandbox-hdp.hortonworks.com:8020/tmp/temp1215132680/tmp1971392817,

Input(s):
Successfully read 194 records (15240 bytes) from: "/user/bibek/flag_project/flag.data"

Output(s):
Successfully stored 6 records (47 bytes) in: "hdfs://sandbox-hdp.hortonworks.com:8020/tmp/
temp1215132680/tmp1971392817"

2021-11-30 04:05:03,647 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat
 - Total input files to process : 1
2021-11-30 04:05:03,647 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.Ma
pRedUtil - Total input paths to process : 1
(1,31)
(2,17)
(3,35)
(4,52)
(5,39)
(6,20)
grunt>
```

*Group E - Notepad

File Edit Format View Help

```
Data Technology Solutions
       Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula
A. Count number of countries based on
   landmass
landmass :
1=N.America, 2=S.America, 3=Europe,
4=Africa, 4=Asia, 6=Oceania
```

Ln 7, Col 17 | 120% | Windows (CRLF) | UTF-8

8:44 AM
01-Dec-21

root@sandbox-hdp:~

```
grunt> foreach_data = FOREACH flag_data GENERATE name,$7+$8;
grunt> grouped_data = GROUP foreach_data All;
grunt> bar_data = FOREACH grouped_data {
>> ord = ORDER foreach_data BY $1 DESC;
>> top = LIMIT ord 5;
>> GENERATE FLATTEN(top);
>> };
grunt> DUMP bar_data;
```

*Group E - Notepad

File Edit Format View Help

```
Data Technology Solutions
       Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula
B. Find out top 5 countries with sum
   of bars and strips in a flag
 $7 & $8 represents bars & strips
 respectively.
```

Ln 16, Col 1 | 120% | Windows (CRLF) | UTF-8

8:47 AM
01-Dec-21

root@sandbox-hdp:~

```
tesRead FileBytesWritten  HdfsBytesRead HdfsBytesWritten Alias  Feature Outputs
scope-149          1          1            194                    0            194
     64             2261          15240                 0 flag_data,foreach_data,grouped_da
ta
scope-150          1          1              0                  194              5
   2261                 0              0                 76 bar_data,grouped_data,ord,top  GR
OUP_BY  hdfs://sandbox-hdp.hortonworks.com:8020/tmp/temp1215132680/tmp143950724,

Input(s):
Successfully read 194 records (15240 bytes) from: "/user/bibek/flag_project/flag.data"

Output(s):
Successfully stored 5 records (76 bytes) in: "hdfs://sandbox-hdp.hortonworks.com:8020/tmp/
temp1215132680/tmp143950724"

2021-11-30 04:12:44,119 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat
 - Total input files to process : 1
2021-11-30 04:12:44,119 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.Ma
pRedUtil - Total input paths to process : 1
(Malaysia,14)
(USA,13)
(Liberia,11)
(Greece,9)
(Uruguay,9)
grunt>
```

*Group E - Notepad

File Edit Format View Help

```
Data Technology Solutions
       Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula
B. Find out top 5 countries with sum
   of bars and strips in a flag
 $7 & $8 represents bars & strips
 respectively.
```

Ln 21, Col 18 | 120% | Windows (CRLF) | UTF-8

8:51 AM
01-Dec-21

root@sandbox-hdp:~

```
grunt> filter_icon_data = FILTER flag_data BY icon == 1;
grunt> group_icon_data = GROUP filter_icon_data BY icon;
grunt> count_icon_data = FOREACH group_icon_data GENERATE group, COUNT(filter_icon_data);
grunt> DUMP count_icon_data;
```

*Group E - Notepad

File Edit Format View Help

```
Data Technology Solutions
       Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

C. Count of countries with icon
 icon: 1 if an inanimate image present
 (e.g., a boat), otherwise 0
```

Ln 44, Col 1 | 120% | Windows (CRLF) | UTF-8

8:55 AM
01-Dec-21

root@sandbox-hdp:~

```
Tez vertex scope-259

Vertex Stats:
VertexId Parallelism TotalTasks   InputRecords   ReduceInputRecords  OutputRecords  FileBy
tesRead FileBytesWritten  HdfsBytesRead HdfsBytesWritten Alias  Feature Outputs
scope-258          1          1            194                    0             49
     64               66          15240                0 count_icon_data,filter_icon_data,
flag_data,group_icon_data
scope-259          1          1              0                    1              1
     66                0             0                7 count_icon_data        GROUP_BY h
dfs://sandbox-hdp.hortonworks.com:8020/tmp/temp1215132680/tmp-925625034,

Input(s):
Successfully read 194 records (15240 bytes) from: "/user/bibek/flag_project/flag.data"

Output(s):
Successfully stored 1 records (7 bytes) in: "hdfs://sandbox-hdp.hortonworks.com:8020/tmp/t
emp1215132680/tmp-925625034"

2021-11-30 04:18:14,010 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat
 - Total input files to process : 1
2021-11-30 04:18:14,010 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.Ma
pRedUtil - Total input paths to process : 1
(1,49)
grunt>
```

*Group E - Notepad

File Edit Format View Help

```
Data Technology Solutions
       Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

C. Count of countries with icon
 icon: 1 if an inanimate image present
 (e.g., a boat), otherwise 0
```

Ln 40, Col 1 | 120% | Windows (CRLF) | UTF-8

8:58 AM
01-Dec-21

root@sandbox-hdp:~

```
grunt> filtered_flag_data = FILTER flag_data BY $28 == $29;
grunt> group_flag_data = GROUP filtered_flag_data All;
grunt> count_flag_data = FOREACH group_flag_data GENERATE group, COUNT(filtered_flag_data)
;
grunt> DUMP count_flag_data;
```

*Group E - Notepad

File Edit Format View Help

```
Data Technology Solutions
        Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

D. Count of countries which have same
  top left and top right color in flag
```

Ln 50, Col 15 | 120% | Windows (CRLF) | UTF-8

1920x1080, 1.709fps, H264 | 00:00:06 | 0.5Mbps

Sort by Date

BANDICUT

Play | Edit | Upload | Delete

Purchase the Bandicam+Bandicut Package! - UP TO 29% OFF

9:00 AM
01-Dec-21

root@sandbox-hdp:~

```
Tez vertex scope-386

Vertex Stats:
VertexId Parallelism TotalTasks   InputRecords   ReduceInputRecords  OutputRecords  FileBy
tesRead FileBytesWritten  HdfsBytesRead HdfsBytesWritten Alias  Feature Outputs
scope-385          1          1           194                   0             76
     64               66          15240                 0 count_flag_data,filtered_flag_dat
a,flag_data,group_flag_data
scope-386          1          1             0                   1              1
     66                0             0                12 count_flag_data        GROUP_BY h
dfs://sandbox-hdp.hortonworks.com:8020/tmp/temp1215132680/tmp-926459893,

Input(s):
Successfully read 194 records (15240 bytes) from: "/user/bibek/flag_project/flag.data"

Output(s):
Successfully stored 1 records (12 bytes) in: "hdfs://sandbox-hdp.hortonworks.com:8020/tmp/
temp1215132680/tmp-926459893"

2021-11-30 04:23:14,719 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat
 - Total input files to process : 1
2021-11-30 04:23:14,719 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.Ma
pRedUtil - Total input paths to process : 1
(all,76)
grunt>
```

*Group E - Notepad

File Edit Format View Help

```
Data Technology Solutions
       Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

D. Count of countries which have same
  top left and top right color in flag
$28 = topleft
$29 = botright
```

Ln 57, Col 15 | 120% | Windows (CRLF) | UTF-8

9:03 AM
01-Dec-21

root@sandbox-hdp:~

```
grunt>
grunt>
grunt>
grunt>
grunt>
grunt>
grunt>
grunt> group_zone_data = GROUP flag_data by zone;
grunt> count_zone_data = FOREACH group_zone_data GENERATE group,COUNT(flag_data);
grunt> DUMP count_zone_data;
2021-11-30 04:26:15,415 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig featu
res used in the script: GROUP_BY
2021-11-30 04:26:15,425 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.sch
ematuple] was not set... will not generate code.
2021-11-30 04:26:15,425 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanO
ptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConst
ParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, Partition
FilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFi
lter, StreamTypeCastInserter]}
2021-11-30 04:26:15,438 [main] INFO  org.apache.pig.backend.hadoop.executionengine.tez.Tez
Launcher - Tez staging directory is /tmp/root/staging and resources directory is /tmp/temp
2026703534
2021-11-30 04:26:15,438 [main] INFO  org.apache.pig.backend.hadoop.executionengine.tez.pla
n.TezCompiler - File concatenation threshold: 100 optimistic? false
2021-11-30 04:26:15,440 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.Co
```

*Group E - Notepad

File Edit Format View Help

```
Data Technology Solutions
       Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula
E. Count number of countries based on
   zone
zone: Geographic quadrant, based on
  Greenwich and the Equator;
  1=NE, 2=SE, 3=SW, 4=NW
```

Ln 68, Col 13 | 120% | Windows (CRLF) | UTF-8

9:05 AM
01-Dec-21

root@sandbox-hdp:~

```
VertexId Parallelism TotalTasks   InputRecords   ReduceInputRecords  OutputRecords  FileBy
tesRead FileBytesWritten  HdfsBytesRead HdfsBytesWritten Alias  Feature Outputs
scope-508          1          1            194                    0            194
     64               93          15240                0 count_zone_data,flag_data,group_z
one_data
scope-509          1          1              0                    4              4
     93                0              0               31 count_zone_data        GROUP_BY h
dfs://sandbox-hdp.hortonworks.com:8020/tmp/temp1215132680/tmp-432907952,

Input(s):
Successfully read 194 records (15240 bytes) from: "/user/bibek/flag_project/flag.data"

Output(s):
Successfully stored 4 records (31 bytes) in: "hdfs://sandbox-hdp.hortonworks.com:8020/tmp/
temp1215132680/tmp-432907952"

2021-11-30 04:26:21,604 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat
 - Total input files to process : 1
2021-11-30 04:26:21,604 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.Ma
pRedUtil - Total input paths to process : 1
(1,91)
(2,29)
(3,16)
(4,58)
grunt>
```

*Group E - Notepad

File Edit Format View Help

```
Data Technology Solutions
       Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula
E. Count number of countries based on
   zone
zone: Geographic quadrant, based on
  Greenwich and the Equator;
  1=NE, 2=SE, 3=SW, 4=NW
```

Ln 68, Col 17 | 120% | Windows (CRLF) | UTF-8

1920x1080, 1.811fps, H264 | 00:00:45 | 0.4Mbps

Sort by Date

BANDICUT

Play | Edit | Upload | Delete

Purchase the Bandicam+Bandicut Package! - UP TO 29% OFF

9:07 AM
01-Dec-21

root@sandbox-hdp:~

```
grunt>
grunt>
grunt> filter_SA_data = FILTER flag_data BY landmass == 2;
grunt> group_SA_data = GROUP filter_SA_data All;
grunt> least_population_data = FOREACH group_SA_data {
>> ord = ORDER filter_SA_data BY $4 ASC;
>> top = LIMIT ord 1;
>> GENERATE FLATTEN(top);
>> };
grunt> least_population_country = FOREACH least_population_data GENERATE name,population;
grunt> DUMP least_population_country;
2021-11-30 04:38:42,895 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig featu
res used in the script: GROUP_BY,FILTER
2021-11-30 04:38:42,907 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.sch
ematuple] was not set... will not generate code.
2021-11-30 04:38:42,908 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanO
ptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConst
ParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, Partition
FilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFi
lter, StreamTypeCastInserter]}
2021-11-30 04:38:42,921 [main] INFO  org.apache.pig.backend.hadoop.executionengine.tez.Tez
Launcher - Tez staging directory is /tmp/root/staging and resources directory is /tmp/temp
2026703534
2021-11-30 04:38:42,921 [main] INFO  org.apache.pig.backend.hadoop.executionengine.tez.pla
n.TezCompiler - File concatenation threshold: 100 optimistic? false
```

*Group E - Notepad

File Edit Format View Help

```
Data Technology Solutions
       Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula
G. Find out least populated country
   in S.America landmass
 landmass 2=S.America
french-Guiana
```

Ln 101, Col 14 | 120% | Windows (CRLF) | UTF-8

9:22 AM
01-Dec-21

root@sandbox-hdp:~

```
Vertex Stats:
VertexId Parallelism TotalTasks   InputRecords   ReduceInputRecords  OutputRecords  FileBy
tesRead FileBytesWritten  HdfsBytesRead HdfsBytesWritten Alias  Feature Outputs
scope-755          1          1            194                   0             17
     64              882          15240                0 filter_SA_data,flag_data,group_SA
_data
scope-756          1          1              0                  17              1
    882                0              0               24 least_population_country,least_po
pulation_data,top       GROUP_BY        hdfs://sandbox-hdp.hortonworks.com:8020/tmp/temp12
15132680/tmp1674816013,

Input(s):
Successfully read 194 records (15240 bytes) from: "/user/bibek/flag_project/flag.data"

Output(s):
Successfully stored 1 records (24 bytes) in: "hdfs://sandbox-hdp.hortonworks.com:8020/tmp/
temp1215132680/tmp1674816013"

2021-11-30 04:41:53,164 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat
 - Total input files to process : 1
2021-11-30 04:41:53,165 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.Ma
pRedUtil - Total input paths to process : 1
(French-Guiana,0)
grunt>
grunt>
```

*Group E - Notepad

File Edit Format View Help

```
Data Technology Solutions
       Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula
G. Find out least populated country
   in S.America landmass
 landmass 2=S.America
 $4 = population
french-Guiana
```

Ln 97, Col 9 | 120% | Windows (CRLF) | UTF-8

```
root@sandbox-hdp:~
grunt>
grunt>
grunt>
grunt>
grunt>
grunt>
grunt>
grunt> foreach_data = FOREACH flag_data GENERATE name,language;
grunt> group_data = GROUP foreach_data BY language;
grunt> count_data = FOREACH group_data GENERATE group,(int)COUNT(foreach_data);
grunt> group_count_data = GROUP count_data All;
grunt> largest_speaking_language = FOREACH group_count_data {
>> ord = ORDER count_data BY $1 DESC;
>> top = LIMIT ord 1;
>> GENERATE FLATTEN(top);
>> };
grunt> DUMP largest_speaking_language;
2021-11-30 04:53:31,611 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig featu
res used in the script: GROUP_BY
2021-11-30 04:53:31,621 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.sch
ematuple] was not set... will not generate code.
2021-11-30 04:53:31,621 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanO
ptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConst
ParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, Partition
FilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFi
```

*Group E - Notepad

File Edit Format View Help

```
Data Technology Solutions

        Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

H. Find out largest speaking language
   among all countries

10 is others.
```

Ln 106, Col 1 | 120% | Windows (CRLF) | UTF-8

9:36 AM
01-Dec-21

PuTTY (inactive)

```
    202             126             0               0 count_data,group_count_data    GR
OUP_BY
scope-793          1          1             0                  10             1
    126               0             0               8 group_count_data,largest_speaking
_language,ord,top       GROUP_BY        hdfs://sandbox-hdp.hortonworks.com:8020/tmp/temp12
15132680/tmp-1613198681,

Input(s):
Successfully read 194 records (15240 bytes) from: "/user/bibek/flag_project/flag.data"

Output(s):
Successfully stored 1 records (8 bytes) in: "hdfs://sandbox-hdp.hortonworks.com:8020/tmp/t
emp1215132680/tmp-1613198681"

2021-11-30 04:53:42,768 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat
 - Total input files to process : 1
2021-11-30 04:53:42,768 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.Ma
pRedUtil - Total input paths to process : 1
(10,46)
grunt>
grunt>
grunt>
grunt>
grunt>
grunt>
```

*Group E - Notepad

File Edit Format View Help

```
Data Technology Solutions

       Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

H. Find out largest speaking language
   among all countries

10 is others.
```

Ln 122, Col 1 | 120% | Windows (CRLF) | UTF-8

10:20 AM
01-Dec-21

root@sandbox-hdp:~

```
Vertex Stats:
VertexId Parallelism TotalTasks   InputRecords   ReduceInputRecords  OutputRecords  FileBy
tesRead FileBytesWritten  HdfsBytesRead HdfsBytesWritten Alias  Feature Outputs
scope-940          1          1            194                   0            194
     64              85          15240               0 common_color_data,flag_data,group
ed_data
scope-941          1          1              0                   1              1
     85               0              0               26 common_color_data      GROUP_BY h
dfs://sandbox-hdp.hortonworks.com:8020/tmp/temp1215132680/tmp186433709,

Input(s):
Successfully read 194 records (15240 bytes) from: "/user/bibek/flag_project/flag.data"

Output(s):
Successfully stored 1 records (26 bytes) in: "hdfs://sandbox-hdp.hortonworks.com:8020/tmp/
temp1215132680/tmp186433709"

2021-11-30 05:00:59,748 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat
 - Total input files to process : 1
2021-11-30 05:00:59,748 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.Ma
pRedUtil - Total input paths to process : 1
(all,153,91,99,91,146,52,26)
grunt>
grunt>
```

*Group E - Notepad

File Edit Format View Help

```
Data Technology Solutions
       Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

I. Find most common color among flags
 from all countries
 red: 153
 green: 91
 blue: 99
gold: 91
white: 146
black:52
orange: 26
```

Ln 133, Col 10 | 120% | Windows (CRLF) | UTF-8

root@sandbox-hdp:~

```
grunt> grouped_circle_data = GROUP flag_data All;
grunt> sum_circle = FOREACH grouped_circle_data GENERATE group,SUM(flag_data.circles);
grunt> DUMP sum_circle;
2021-11-30 05:05:57,992 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig featu
res used in the script: GROUP_BY
2021-11-30 05:05:58,001 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.sch
ematuple] was not set... will not generate code.
2021-11-30 05:05:58,001 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanO
ptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConst
ParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, Partition
FilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFi
lter, StreamTypeCastInserter]}
2021-11-30 05:05:58,012 [main] INFO  org.apache.pig.backend.hadoop.executionengine.tez.Tez
Launcher - Tez staging directory is /tmp/root/staging and resources directory is /tmp/temp
2026703534
2021-11-30 05:05:58,013 [main] INFO  org.apache.pig.backend.hadoop.executionengine.tez.pla
n.TezCompiler - File concatenation threshold: 100 optimistic? false
2021-11-30 05:05:58,014 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.Co
mbinerOptimizerUtil - Choosing to move algebraic foreach to combiner
2021-11-30 05:05:58,025 [main] INFO  org.apache.pig.builtin.PigStorage - Using PigTextInpu
tFormat
2021-11-30 05:05:58,027 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat
 - Total input files to process : 1
2021-11-30 05:05:58,027 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.Ma
pRedUtil - Total input paths to process : 1
```

*Group E - Notepad

File Edit Format View Help

```
Data Technology Solutions
       Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula


J. Sum of all circles present in all
   country flags.
  33
```

Ln 152, Col 1 | 120% | Windows (CRLF) | UTF-8

1920x1080, 2.899fps, H264 | 00:00:49 | 0.4Mbps

Sort by Date

BANDICUT

Play | Edit | Upload | Delete

Purchase the Bandicam+Bandicut Package! - UP TO 29% OFF

9:48 AM
01-Dec-21

root@sandbox-hdp:~

```
Tez vertex scope-1227

Vertex Stats:
VertexId Parallelism TotalTasks   InputRecords   ReduceInputRecords  OutputRecords  FileBy
tesRead FileBytesWritten  HdfsBytesRead HdfsBytesWritten Alias  Feature Outputs
scope-1226         1          1            194                    0            194
     64               66          15240                0 flag_data,grouped_circle_data,sum
_circle
scope-1227         1          1              0                    1              1
     66                0              0                12 sum_circle     GROUP_BY        hd
fs://sandbox-hdp.hortonworks.com:8020/tmp/temp1215132680/tmp-1636563055,

Input(s):
Successfully read 194 records (15240 bytes) from: "/user/bibek/flag_project/flag.data"

Output(s):
Successfully stored 1 records (12 bytes) in: "hdfs://sandbox-hdp.hortonworks.com:8020/tmp/
temp1215132680/tmp-1636563055"

2021-11-30 05:06:02,136 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat
 - Total input files to process : 1
2021-11-30 05:06:02,136 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.Ma
pRedUtil - Total input paths to process : 1
(all,33)
grunt>
```

*Group E - Notepad

File Edit Format View Help

```
Data Technology Solutions
       Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula


J. Sum of all circles present in all
   country flags.
  33
```

Ln 152, Col 1 | 120% | Windows (CRLF) | UTF-8

1920x1080, 2.899fps, H264 | 00:00:49 | 0.4Mbps

Sort by Date

BANDICUT

Play | Edit | Upload | Delete

Purchase the Bandicam+Bandicut Package! - UP TO 29% OFF

9:48 AM
01-Dec-21