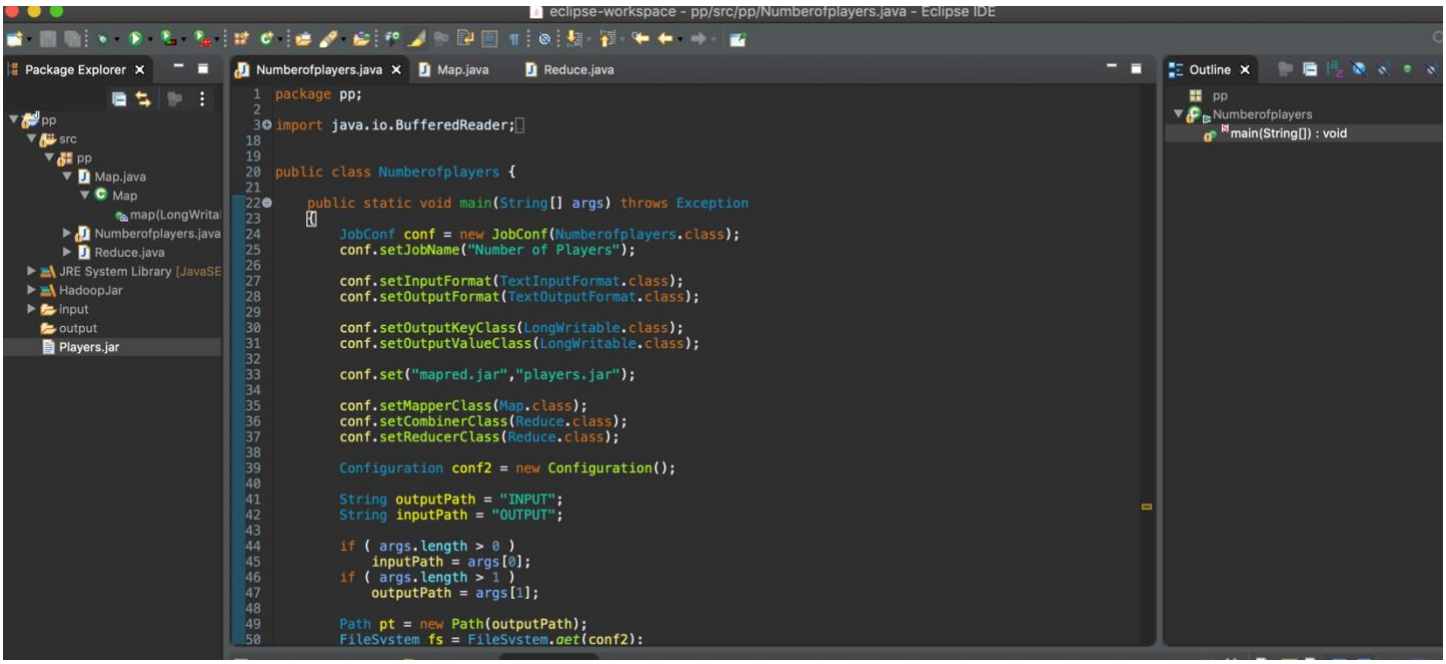
Homework-4

Apache Hadoop – Java

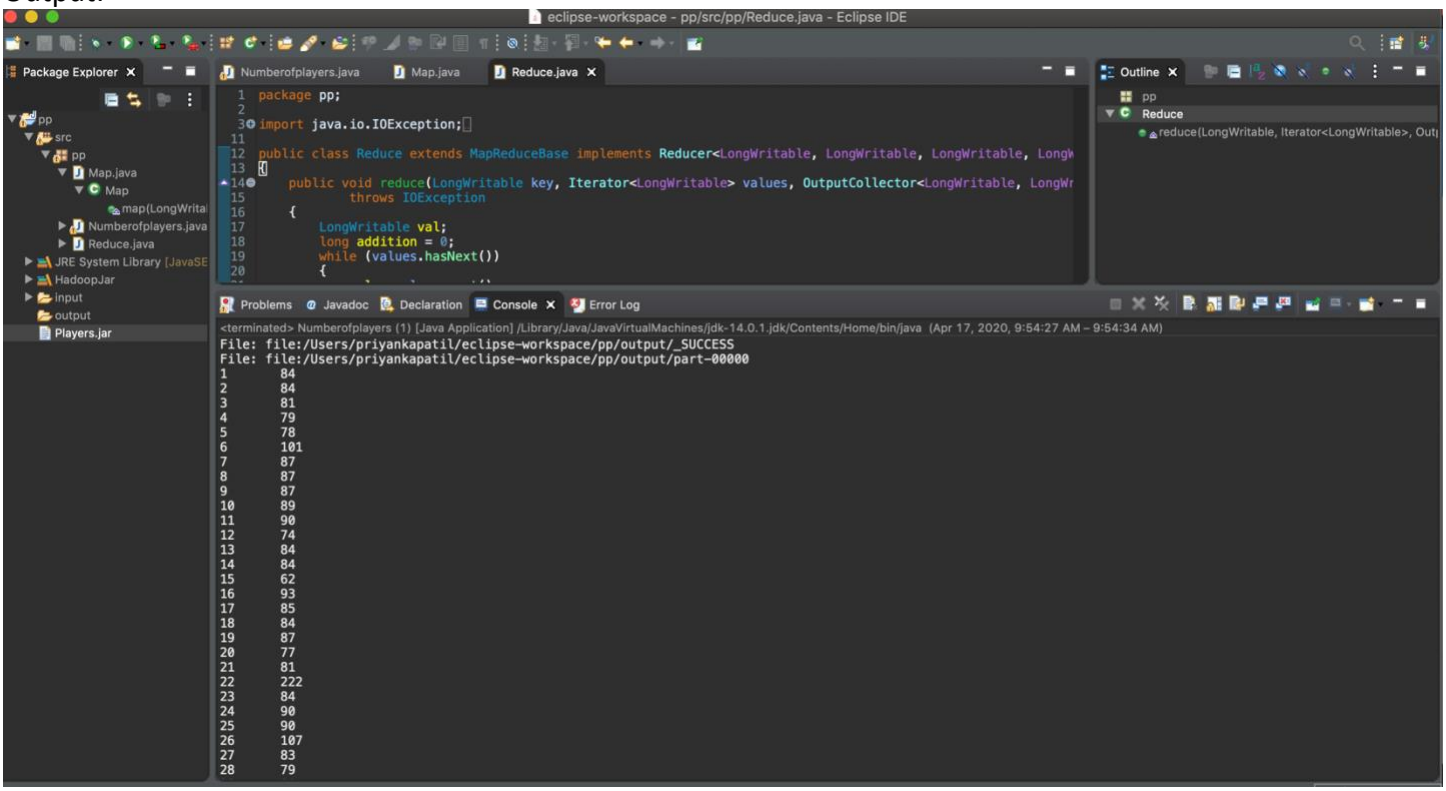
Question-1: (3 points)

Write a Hadoop MapReduce program in Java that will calculate the number of players per game.

Code:



Output:



```
<terminated> Numberofplayers (1) [Java Application] /Library/Java/JavaVirtualMachines/jdk-14.0.1.jdk/Contents/Home/bin/java  (Apr 17, 2020, 9:54:27 AM – 9:54:34 AM)
73      88
74      96
75      92
76      98
77      219
78      99
79      77
80      86
81      86
82      90
83      214
84      95
85      226
86      226
87      96
88      92
89      74
90      92
91      78
92      217
93      94
94      90
95      80
96      75
97      81
98      76
99      99
100     87
Number of Players per game: 100
```

Apache Hive

Question-2: (1 Point)

Implement Word Search Program in Hive

```
hive> use hivetest1;
OK
Time taken: 0.026 seconds
hive> create table constitution(Description String);
OK
Time taken: 0.18 seconds
hive> describe constitution;
OK
description             string
Time taken: 0.055 seconds, Fetched: 1 row(s)
hive> LOAD DATA INPATH '/user/ppatil3/hive1/constitution.txt' into table constitution;
Loading data to table hivetest1.constitution
OK
Time taken: 0.629 seconds
```

```
[ppatil3@login ~] hadoop fs -put constitution.txt '/user/ppatil3/hive1' ;
[ppatil3@login ~] hadoop fs -put declaration.txt '/user/ppatil3/hive1' ;
[ppatil3@login ~]
```

Output:

```
hive> LOAD DATA INPATH '/user/ppatil3/hive1/declaration.txt' into table constitution;
Loading data to table hivetest1.constitution
OK
Time taken: 0.569 seconds
hive>  select word,count(lower(word)) from constitution c lateral view explode(split(lower(c.Description), ' ')) tb as word where lower(word) = 'the' group by word;
Query ID = ppatil3_20200417080335_d68ee185-cd70-4fbb-a2c3-a920a670395c
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
20/04/17 08:03:35 INFO client.RMProxy: Connecting to ResourceManager at worker2.hdp-internal/10.3.0.2:8032
20/04/17 08:03:35 INFO client.RMProxy: Connecting to ResourceManager at worker2.hdp-internal/10.3.0.2:8032
Starting Job = job_1586968973396_0598, Tracking URL = http://worker2.hdp-internal:8088/proxy/application_1586968973396_0598/
Kill Command = /opt/cloudera/parcels/CDH-6.2.0-1.cdh6.2.0.p0.967373/lib/hadoop/bin/hadoop job  -kill job_1586968973396_0598
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2020-04-17 08:03:42,456 Stage-1 map = 0%,  reduce = 0%
2020-04-17 08:03:48,652 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 12.02 sec
2020-04-17 08:03:52,800 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 15.74 sec
MapReduce Total cumulative CPU time: 15 seconds 740 msec
Ended Job = job_1586968973396_0598
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 1   Cumulative CPU: 15.74 sec   HDFS Read: 71945 HDFS Write: 107 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 15 seconds 740 msec
OK
the     805
Time taken: 19.522 seconds, Fetched: 1 row(s)
```

Question-3: (3 points – 15M/5)

Load the Data into Hive Tables. The files games, players and player_games are provided on Blackboard

```
hive> create external table games1_p(gid int,gname string, pubname string, releasedate string, rating double)
    > COMMENT 'games' ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n' STORED AS TEXTFILE;
OK
Time taken: 0.374 seconds
hive> CREATE EXTERNAL TABLE players1_p(pid INT, fname STRING, lname STRING, bdate STRING, gender STRING, imagePath STRING)
    > COMMENT 'players' ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n' STORED AS TEXTFILE;
OK
Time taken: 0.216 seconds
hive> CREATE EXTERNAL TABLE playergames1_p (pid INT, gid INT, score INT)
    > COMMENT 'playergames' ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n' STORED AS TEXTFILE;
OK
Time taken: 0.084 seconds
hive> describe games1_p;
OK
gid                     int
gname                   string
pubname                 string
releasedate             string
rating                  double
Time taken: 0.059 seconds, Fetched: 5 row(s)
hive> describe players1_p;
OK
pid                     int
fname                   string
lname                   string
bdate                   string
gender                  string
imagepath               string
Time taken: 0.057 seconds, Fetched: 6 row(s)
hive> describe playergames_p;
FAILED: SemanticException [Error 10001]: Table not found playergames_p
hive> describe playergames1_p;
OK
pid                     int
gid                     int
score                   int
Time taken: 0.055 seconds, Fetched: 3 row(s)
```

```
[ppatil3@login ~] ls
Bigdata1          declaration.txt  pig_1587112391231.log  pig_1587115534103.log  pig_1587116984716.log  pig_1587118923045.log  u.data
bin               games.txt        pig_1587114733850.log  pig_1587116552407.log  pig_1587117343789.log  player_games.txt
constitution.txt  index.html       pig_1587115355254.log  pig_1587116680821.log  pig_1587118623747.log  players.txt
[ppatil3@login ~] hadoop fs -put games.txt '/user/ppatil3/hive1';
put: `/user/ppatil3/hive1/games.txt': File exists
[ppatil3@login ~] hadoop fs -put players.txt '/user/ppatil3/hive1' ;
put: `/user/ppatil3/hive1/players.txt': File exists
[ppatil3@login ~] hadoop fs -put player_games.txt '/user/ppatil3/hive1' ;
put: `/user/ppatil3/hive1/player_games.txt': File exists
[ppatil3@login ~]
```

```
hive> LOAD DATA INPATH '/user/ppatil3/hive1/games.txt' into table games1_p;
Loading data to table hivetest1.games1_p
OK
Time taken: 0.866 seconds
hive> LOAD DATA INPATH '/user/ppatil3/hive1/players.txt' into table players1_p;
Loading data to table hivetest1.players1_p
OK
Time taken: 0.624 seconds
hive> LOAD DATA INPATH '/user/ppatil3/hive1/player_games.txt' into table playergames1_p;
Loading data to table hivetest1.playergames1_p
OK
Time taken: 0.653 seconds
```

Task #1 (3 marks) - Players 18 and over

Write a HiveQL query that lists only the players that are 18 and over. The output should be sorted by age ascending. You can do this question without a UDF if you look at the date function support of Hive.

```
hive> use hivetest1;
OK
Time taken: 0.435 seconds
hive> select * from players1_p where DATEDIFF(current_date,bdate)>18 order by bdate asc;
Query ID = ppatil3_20200417081908_bebecd21-b357-41d2-87fe-3c80cf59ed3d
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
20/04/17 08:19:09 INFO client.RMProxy: Connecting to ResourceManager at worker2.hdp-internal/10.3.0.2:8032
20/04/17 08:19:09 INFO client.RMProxy: Connecting to ResourceManager at worker2.hdp-internal/10.3.0.2:8032
Starting Job = job_1586968973396_0602, Tracking URL = http://worker2.hdp-internal:8088/proxy/application_1586968973396_0602/
Kill Command = /opt/cloudera/parcels/CDH-6.2.0-1.cdh6.2.0.p0.967373/lib/hadoop/bin/hadoop job  -kill job_1586968973396_0602
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2020-04-17 08:19:16,099 Stage-1 map = 0%,  reduce = 0%
2020-04-17 08:19:21,342 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 5.01 sec
2020-04-17 08:19:26,504 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 8.69 sec
MapReduce Total cumulative CPU time: 8 seconds 690 msec
Ended Job = job_1586968973396_0602
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 8.69 sec   HDFS Read: 69282 HDFS Write: 71203 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 690 msec
OK
```

```
622     Alex    Campbell        1996-01-19      male    /data/images/male/3.jpg
652     Kevin   Scott   1996-01-19      male    /data/images/male/7.jpg
662     Gary    Clark   1996-01-19      male    /data/images/male/7.jpg
664     Matt    Howard  1996-01-19      male    /data/images/male/1.jpg
667     Amy     Wood    1996-01-19      female  /data/images/female/7.jpg
60      Adrienne        Rogers  1996-01-19      female  /data/images/female/5.jpg
674     Jamie   Gonzales        1996-01-19      female  /data/images/female/6.jpg
682     Craig   Myers   1996-01-19      male    /data/images/male/9.jpg
690     Brendon Perry   1996-01-19      male    /data/images/male/7.jpg
54      Logan   Rivera  1996-01-19      male    /data/images/male/4.jpg
692     Trevor  Bryant  1996-01-19      male    /data/images/male/7.jpg
712     Chantel Watson  1996-01-19      female  /data/images/female/4.jpg
714     Mallory Rivera  1996-01-19      female  /data/images/female/1.jpg
716     Jose    White   1996-01-19      male    /data/images/male/9.jpg
720     Drew    Flores  1996-01-19      male    /data/images/male/6.jpg
728     Jack    Sanders 1996-01-19      male    /data/images/male/2.jpg
729     Paul    Bennett 1996-01-19      male    /data/images/male/3.jpg
731     Christine       Powell  1996-01-19      female  /data/images/female/2.jpg
736     Deborah Miller  1996-01-19      female  /data/images/female/6.jpg
758     Lane    Lewis   1996-01-19      male    /data/images/male/8.jpg
759     Jane    Evans   1996-01-19      female  /data/images/female/1.jpg
779     Deborah Howard  1996-01-19      female  /data/images/female/9.jpg
781     Brooke  Cook    1996-01-19      female  /data/images/female/8.jpg
791     Mike    Barnes  1996-01-19      male    /data/images/male/4.jpg
792     Chris   Russell 1996-01-19      male    /data/images/male/1.jpg
824     Katie   Jones   1996-01-19      female  /data/images/female/5.jpg
825     Jaquline        Williams        1996-01-19      female  /data/images/female/8.jpg
833     Julie   Gray    1996-01-19      female  /data/images/female/3.jpg
835     April   Griffin 1996-01-19      female  /data/images/female/3.jpg
847     Andrea  Phillips        1996-01-19      female  /data/images/female/4.jpg
850     Abby    Lewis   1996-01-19      female  /data/images/female/3.jpg
877     Rob     Phillips        1996-01-19      male    /data/images/male/1.jpg
884     Evan    Russell 1996-01-19      male    /data/images/male/7.jpg
891     Jen     Gray    1996-01-19      female  /data/images/female/7.jpg
897     Sarah   Davis   1996-01-19      female  /data/images/female/6.jpg
904     Mary    Young   1996-01-19      female  /data/images/female/6.jpg
922     Chantel Ramirez 1996-01-19      female  /data/images/female/6.jpg
924     Maria   Lee     1996-01-19      female  /data/images/female/9.jpg
925     Deborah Barnes  1996-01-19      female  /data/images/female/5.jpg
12      Steph   Bennett 1996-01-19      female  /data/images/female/8.jpg
965     Vince   Jones   1996-01-19      male    /data/images/male/2.jpg
970     Kevin   Gonzales        1996-01-19      male    /data/images/male/9.jpg
7       Scott   Bell    1996-01-19      male    /data/images/male/4.jpg
5       Sasha   Gray    1996-01-19      female  /data/images/female/2.jpg
987     Lane    Richardson      1996-01-19      male    /data/images/male/8.jpg
2       Sharon  Cleo    1996-01-19      female  /data/images/female/6.jpg
1000    Patrick Jones   1996-01-19      male    /data/images/male/7.jpg
Time taken: 20.242 seconds, Fetched: 1000 row(s)
```

Task #2 (2 marks) - Players in Common

Write a HiveQL query that will output pairs of game ids and the number of players they have in common. For instance, if game X and game Y have 2,000 players in common (play both games), then output X, Y, 2000. The data does not have to be sorted.

priyankapatil — ssh ppatil3@login.hadoop.umbc.edu — 169×49

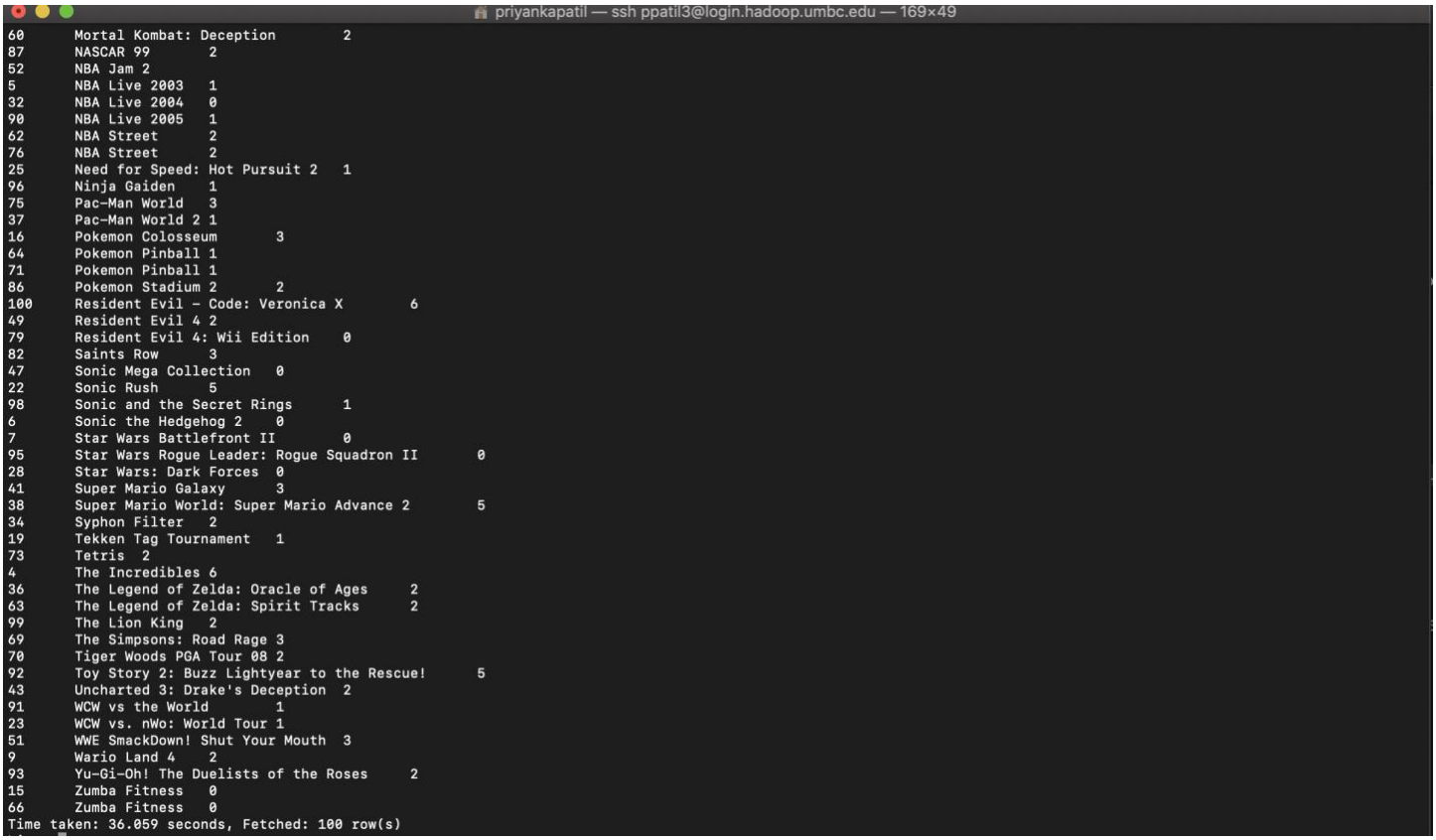
```
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 14.09 sec   HDFS Read: 138693 HDFS Write: 199009 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 14 seconds 90 msec
OK
1       2       5
1       3       7
1       4       8
1       5       3
1       6       4
1       7       6
1       8       5
1       9       5
1       10      6
1       11      3
1       12      11
1       13      5
1       14      3
1       15      5
1       16      12
1       17      8
1       18      4
1       19      8
1       20      9
1       21      10
1       22      14
1       23      8
1       24      11
1       25      6
1       26      9
1       27      8
1       28      9
1       29      8
1       30      19
1       32      7
1       33      2
1       34      10
1       35      9
1       36      6
1       37      15
1       38      16
1       39      9
1       40      4
1       41      11
1       42      6
1       43      7
1       44      17
1       45      13
1       46      9
1       47      12
1       48      10
```

priyankapatil — ssh ppatil3@login.hadoop.umbc.edu — 169×49

```
100     53      5
100     54      22
100     55      8
100     56      7
100     57      7
100     58      11
100     59      9
100     60      11
100     61      4
100     62      7
100     63      7
100     64      9
100     65      6
100     66      4
100     67      12
100     68      4
100     69      9
100     70      9
100     71      9
100     72      7
100     73      8
100     74      7
100     75      6
100     76      6
100     77      34
100     78      8
100     79      6
100     80      6
100     81      9
100     82      3
100     83      16
100     84      9
100     85      18
100     86      23
100     87      12
100     88      10
100     89      6
100     90      7
100     91      9
100     92      22
100     93      12
100     94      14
100     95      10
100     96      4
100     97      5
100     98      8
100     99      10
Time taken: 28.278 seconds, Fetched: 9836 row(s)
```

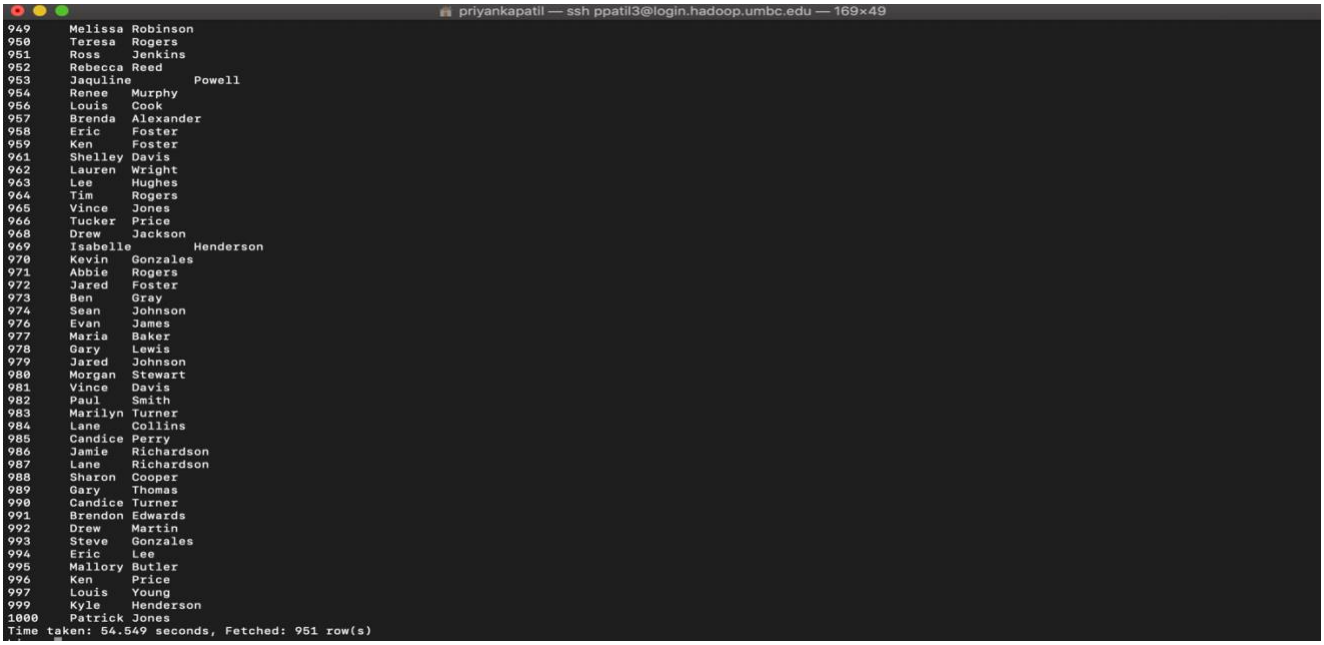
Task #3 (3 marks) - Count Top Players for Each Game

Write a HiveQL query that will list all games along with the count of the number of players in each game with a score over 98,000. If a game does not have a player with a score over 98,000, it should still appear in the output with a count of 0.



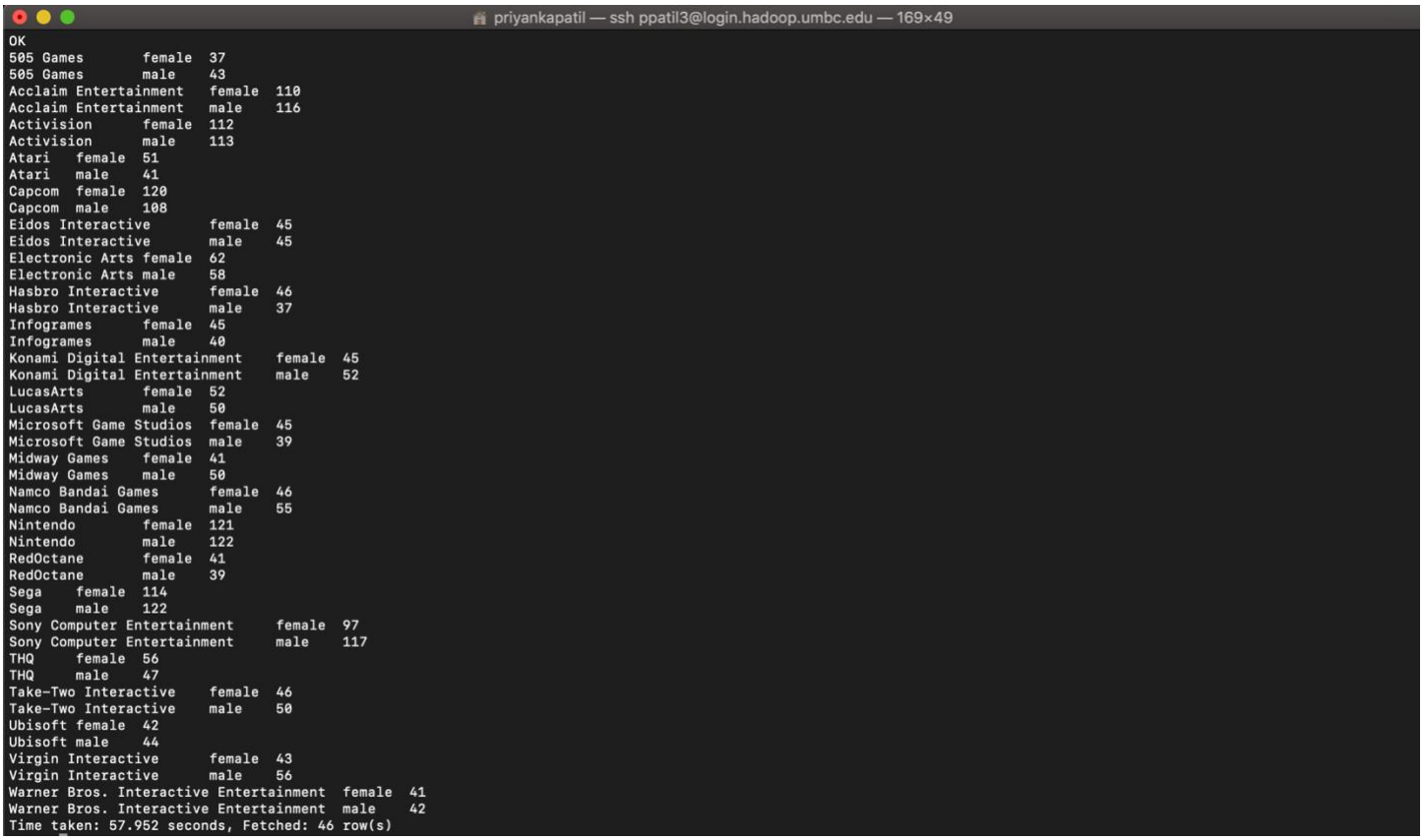
Task #4 (4 marks) - List All Players with Certain Properties

Write a HiveQL query that will list all players that either have a score in some game over 90,000 or play a game published by 'Electronic Arts'.



Task #5 (3 marks) - List Most Popular Games for Each Publisher by Gender

Write a HiveQL query that for each publisher will list two records. The first will be the publisher id, "female", and the maximum number of women that play one of its games. The second row will be the publisher id, "male", and the maximum number of men that play one of its games. Note that the games may be different.



Apache Pig

Question-4: (3 Points – 15/5)

Task #1 (2 marks) - List all Games

Write a Pig script that will list all the game records. There should be 100 records printed.

```
Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime        Alias   Feature Outpu
ts
job_1586968973396_0549  1       0       2       2       2       2       0       0       0       0       C       MAP_ONLY        hdfs://worker2.hdp-internal:8020/tmp/temp570241069/tm
p553840282,

Input(s):
Successfully read 100 records (5858 bytes) from: "hdfs://worker2.hdp-internal:8020/user/ppatil3/hive1/games.txt"

Output(s):
Successfully stored 100 records (6317 bytes) in: "hdfs://worker2.hdp-internal:8020/tmp/temp570241069/tmp553840282"

Counters:
Total records written : 100
Total bytes written : 6317
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
```

```
2020-04-17 06:38:33,728 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2020-04-17 06:38:33,740 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2020-04-17 06:38:33,740 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,Madden NFL 10,Electronic Arts,2009-10-05,0.487)
(2,Medal of Honor Heroes,Electronic Arts,2006-12-29,0.401)
(3,Halo 4,Microsoft Game Studios,2012-07-09,0.828)
(4,The Incredibles,THQ,2004-06-04,0.005)
(5,NBA Live 2003,Electronic Arts,2002-07-28,0.111)
(6,Sonic the Hedgehog 2,Sega,1992-04-06,0.661)
(7,Star Wars Battlefront II,LucasArts,2005-11-13,0.398)
(8,ATV Offroad Fury,Sony Computer Entertainment,2001-05-19,0.907)
(9,Wario Land 4,Nintendo,2001-04-22,0.821)
(10,Mario Party 9,Nintendo,2012-06-27,0.799)
(11,Madden NFL 08,Electronic Arts,2007-10-24,0.563)
(12,Carnival Games,Take-Two Interactive,2007-02-13,0.357)
(13,Jeremy McGrath Supercross 98,Acclaim Entertainment,2003-03-05,0.243)
(14,God of War II,Sony Computer Entertainment,2007-07-14,0.213)
(15,Zumba Fitness,505 Games,2010-12-03,0.208)
(16,Pokemon Colosseum,Nintendo,2003-10-10,0.765)
(17,Kung Fu Panda,Activision,2008-03-20,0.038)
(18,Dr. Mario,Nintendo,1990-11-27,0.520)
(19,Tekken Tag Tournament,Namco Bandai Games,2000-10-20,0.292)
(20,Guitar Hero III: Legends of Rock,Activision,2007-11-22,0.612)
(21,Battlefield 3,Electronic Arts,2011-09-08,0.429)
(22,Sonic Rush,Sega,2005-05-04,0.010)
(23,WCW vs. nWo: World Tour,THQ,1997-03-19,0.838)
(24,Madden NFL 2002,Electronic Arts,2001-06-16,0.034)
(25,Need for Speed: Hot Pursuit 2,Electronic Arts,2002-09-27,0.841)
(26,Medal of Honor,Electronic Arts,1998-06-27,0.251)
(27,Monopoly,Hasbro Interactive,1994-10-10,0.363)
(28,Star Wars: Dark Forces,LucasArts,1992-04-18,0.735)
(29,Golf,Nintendo,1989-03-09,0.880)
(30,Call of Duty: Black Ops,Activision,2010-01-29,0.034)
(31,Carnival Games,Take-Two Interactive,2008-06-26,0.796)
(32,NBA Live 2004,Electronic Arts,2003-09-25,0.063)
(33,Batman: Arkham Asylum,Eidos Interactive,2009-07-24,0.210)
(34,Syphon Filter,Sony Computer Entertainment,1999-03-02,0.837)
(35,Gran Turismo 5,Sony Computer Entertainment,2010-12-16,0.373)
(36,The Legend of Zelda: Oracle of Ages,Nintendo,2001-08-28,0.292)
(37,Pac-Man World 2,Namco Bandai Games,2002-09-21,0.661)
(38,Super Mario World: Super Mario Advance 2,Nintendo,2001-11-20,0.360)
(39,MechAssault,Microsoft Game Studios,2002-05-16,0.753)
(40,Just Dance 3,Ubisoft,2011-12-03,0.709)
(41,Super Mario Galaxy,Nintendo,2007-06-20,0.147)
(42,Metal Gear Solid 4: Guns of the Patriots,Konami Digital Entertainment,2008-02-27,0.647)
(43,Uncharted 3: Drake's Deception,Sony Computer Entertainment,2011-02-21,0.266)
(44,Devil May Cry 3: Dante's Awakening Special Edition,Capcom,2006-11-03,0.101)
(45,Madden NFL 10,Electronic Arts,2009-10-09,0.708)
(46,Grand Theft Auto: Vice City Stories,Take-Two Interactive,2006-03-16,0.163)
(47,Sonic Mega Collection,Infogrames,2002-09-13,0.755)
(48,Dragon Age: Origins,Electronic Arts,2009-05-01,0.815)
(49,Resident Evil 4,Capcom,2005-01-11,0.473)
(50,Kirby's Dream Land,Nintendo,1992-02-29,0.167)
(51,WWE SmackDown! Shut Your Mouth,THQ,2002-08-13,0.653)
(52,NBA Jam,Acclaim Entertainment,1994-07-11,0.936)
```

Task #2 (3 marks) - Find a Game by its Id

Write a Pig script that given some game id (hard-coded constant) will return the game record if found.

```
Input(s):
Successfully read 100 records (5858 bytes) from: "hdfs://worker2.hdp-internal:8020/user/ppatil3/hive1/games.txt"

Output(s):
Successfully stored 1 records (68 bytes) in: "hdfs://worker2.hdp-internal:8020/tmp/temp570241069/tmp-2064050571"

Counters:
Total records written : 1
Total bytes written : 68
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1586968973396_0550


2020-04-17 06:39:31,525 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at worker2.hdp-internal/10.3.0.2:8032
2020-04-17 06:39:31,529 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history se
rver
2020-04-17 06:39:31,547 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at worker2.hdp-internal/10.3.0.2:8032
2020-04-17 06:39:31,550 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history se
rver
2020-04-17 06:39:31,565 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at worker2.hdp-internal/10.3.0.2:8032
2020-04-17 06:39:31,568 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history se
rver
2020-04-17 06:39:31,585 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2020-04-17 06:39:31,585 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system
-metrics-publisher.enabled
2020-04-17 06:39:31,586 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2020-04-17 06:39:31,592 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2020-04-17 06:39:31,592 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(45,Madden NFL 10,Electronic Arts,2009-10-09,0.708)
```

Task #3 (3 marks) - Number of Players per Game

Write a Pig script that will calculate the number of players per game. The output does not have to be sorted.

```
Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime                        A
lias    Feature Outputs
job_1586968973396_0556  1       1       3       3       3       3       2       2       2       2       X,Y,plg GROUP_BY,COMBINER       hdfs://worker2.hdp-internal:8020/tmp/temp-201
6907815/tmp987573200,

Input(s):
Successfully read 10000 records (127554 bytes) from: "hdfs://worker2.hdp-internal:8020/user/ppatil3/hive1/player_games.txt"

Output(s):
Successfully stored 100 records (519 bytes) in: "hdfs://worker2.hdp-internal:8020/tmp/temp-2016907815/tmp987573200"

Counters:
Total records written : 100
Total bytes written : 519
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1586968973396_0556
```

```
(1,84)
(2,84)
(3,81)
(4,79)
(5,78)
(6,101)
(7,87)
(8,87)
(9,87)
(10,89)
(11,90)
(12,74)
(13,84)
(14,84)
(15,62)
(16,93)
(17,85)
(18,84)
(19,87)
(20,77)
(21,81)
(22,222)
(23,84)
(24,90)
(25,90)
(26,107)
(27,83)
(28,79)
(29,90)
(30,213)
(31,13)
(32,98)
(33,90)
(34,98)
(35,102)
(36,83)
(37,101)
(38,236)
(39,71)
(40,82)
(41,104)
(42,88)
(43,88)
(44,228)
(45,97)
(46,96)
(47,85)
(48,98)
(49,86)
(50,70)
(51,95)
(52,88)
```

Task #4 (2 marks) - Given a Game Id - List the Top 10 Scores for the Game

Write a Pig script that will output the top 10 scores in descending order for a given game id.

```
Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime                        A
lias    Feature Outputs
job_1586968973396_0559  1       0       2       2       2       2       0       0       0       0       W,plg   MAP_ONLY
job_1586968973396_0560  1       1       2       2       2       2       3       3       3       3       X       SAMPLER
job_1586968973396_0561  1       1       2       2       2       2       3       3       3       3       X       ORDER_BY,COMBINER
job_1586968973396_0562  1       1       2       2       2       2       2       2       2       2       X,Y             hdfs://worker2.hdp-internal:8020/tmp/temp-2016907815/tmp-1683
539096,

Input(s):
Successfully read 10000 records (127554 bytes) from: "hdfs://worker2.hdp-internal:8020/user/ppatil3/hive1/player_games.txt"

Output(s):
Successfully stored 10 records (97 bytes) in: "hdfs://worker2.hdp-internal:8020/tmp/temp-2016907815/tmp-1683539096"

Counters:
Total records written : 10
Total bytes written : 97
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
```

```
2020-04-17 06:58:39,731 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(678,99006)
(695,98444)
(462,98271)
(9,96876)
(225,93114)
(349,92377)
(291,89110)
(21,87674)
(629,85911)
(76,85308)
grunt>
```

Task #5 (5 marks) - Show Game Breakdown by Gender

For each game, show the total number of players, the total number of female and male players, and the percentage of female and male players.

```
Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime                        A
lias    Feature Outputs
job_1586968973396_0589  2       1       2       2       2       2       2       2       2       2       Af,Bf,X HASH_JOIN
job_1586968973396_0590  1       1       2       2       2       2       2       2       2       2       Y,Z     GROUP_BY,COMBINER       hdfs://worker2.hdp-internal:8020/tmp/temp-162
5104245/tmp286296160,

Input(s):
Successfully read 1000 records from: "hdfs://worker2.hdp-internal:8020/user/ppatil3/hive1/players.txt"
Successfully read 10000 records from: "hdfs://worker2.hdp-internal:8020/user/ppatil3/hive1/player_games.txt"

Output(s):
Successfully stored 200 records (2618 bytes) in: "hdfs://worker2.hdp-internal:8020/tmp/temp-1625104245/tmp286296160"

Counters:
Total records written : 200
Total bytes written : 2618
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
```

```
(1,male,49)
(1,female,35)
(2,male,42)
(2,female,42)
(3,male,39)
(3,female,42)
(4,male,44)
(4,female,35)
(5,male,36)
(5,female,42)
(6,male,46)
(6,female,55)
(7,male,50)
(7,female,37)
(8,male,40)
(8,female,47)
(9,male,52)
(9,female,35)
(10,male,50)
(10,female,39)
(11,male,39)
(11,female,51)
(12,male,39)
(12,female,35)
(13,male,41)
(13,female,43)
(14,male,44)
(14,female,40)
(15,male,33)
(15,female,29)
(16,male,49)
(16,female,44)
(17,male,33)
(17,female,52)
(18,male,50)
(18,female,34)
(19,male,43)
(19,female,44)
(20,male,38)
(20,female,39)
(21,male,37)
(21,female,44)
(22,male,108)
```