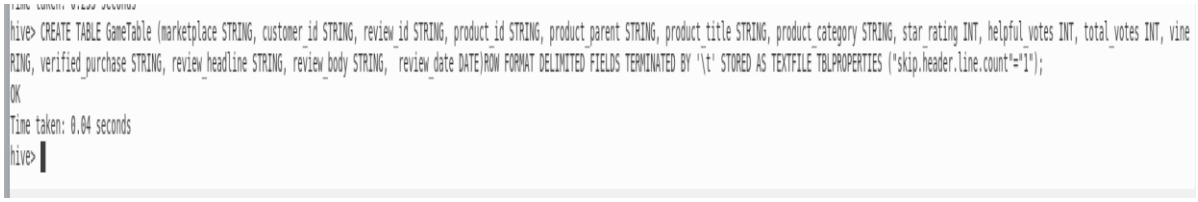
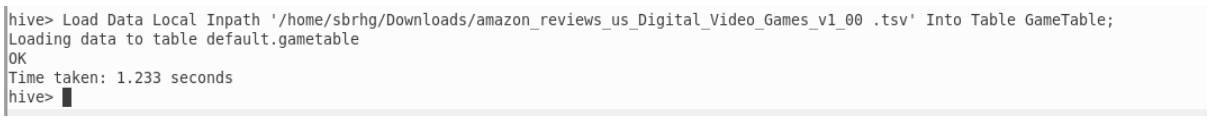
STUDENT-1

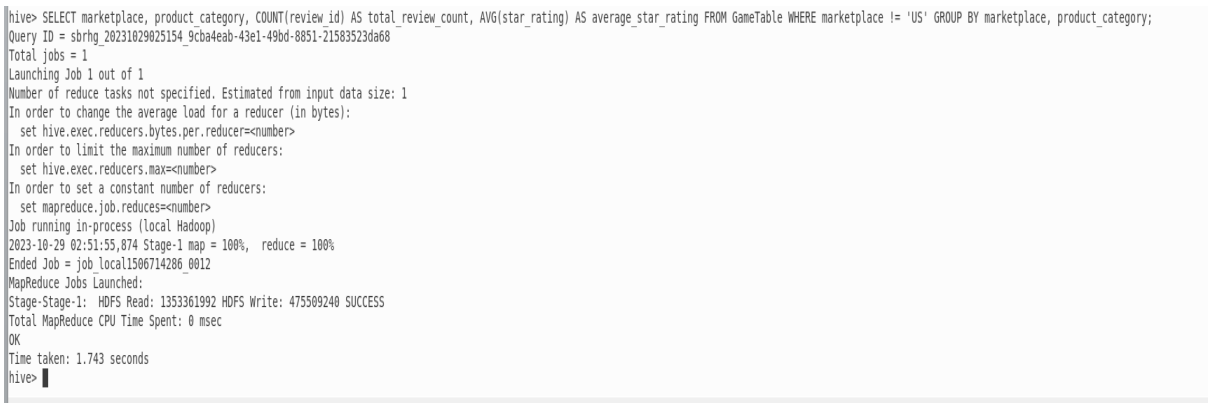
CREATE TABLE GameTable (marketplace STRING, customer_id STRING, review_id STRING, product_id STRING, product_parent STRING, product_title STRING, product_category STRING, star_rating INT, helpful_votes INT, total_votes INT, vine STRING, verified_purchase STRING, review_headline STRING, review_body STRING, review_date DATE)ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS TEXTFILE TBLPROPERTIES ("skip.header.line.count"="1");



Load Data Local Inpath '/home/sbrhg/Downloads/amazon_reviews_us_Digital_Video_Games_v1_00 .tsv' Into Table GameTable;

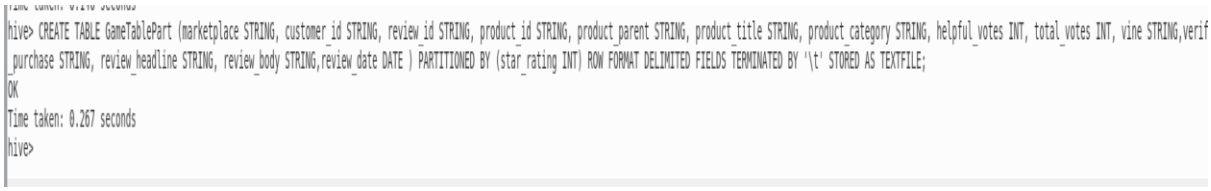


SELECT marketplace, product_category, COUNT(review_id) AS total_review_count, AVG(star_rating) AS average_star_rating FROM GameTable WHERE marketplace != 'US' GROUP BY marketplace, product_category;

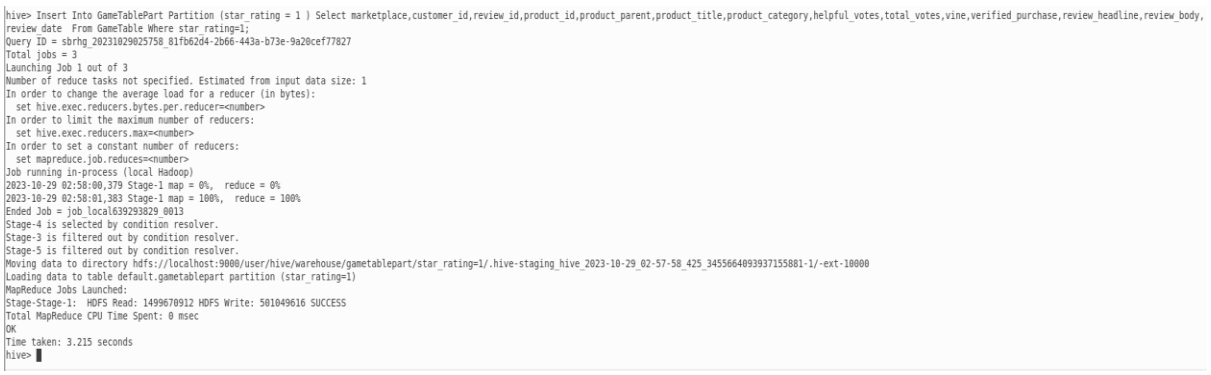


STUDENT-2

CREATE TABLE GameTablePart (marketplace STRING, customer_id STRING, review_id STRING, product_id STRING, product_parent STRING, product_title STRING, product_category STRING, helpful_votes INT, total_votes INT, vine STRING,verified_purchase STRING, review_headline STRING, review_body STRING,review_date DATE) PARTITIONED BY (star_rating INT) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS TEXTFILE;



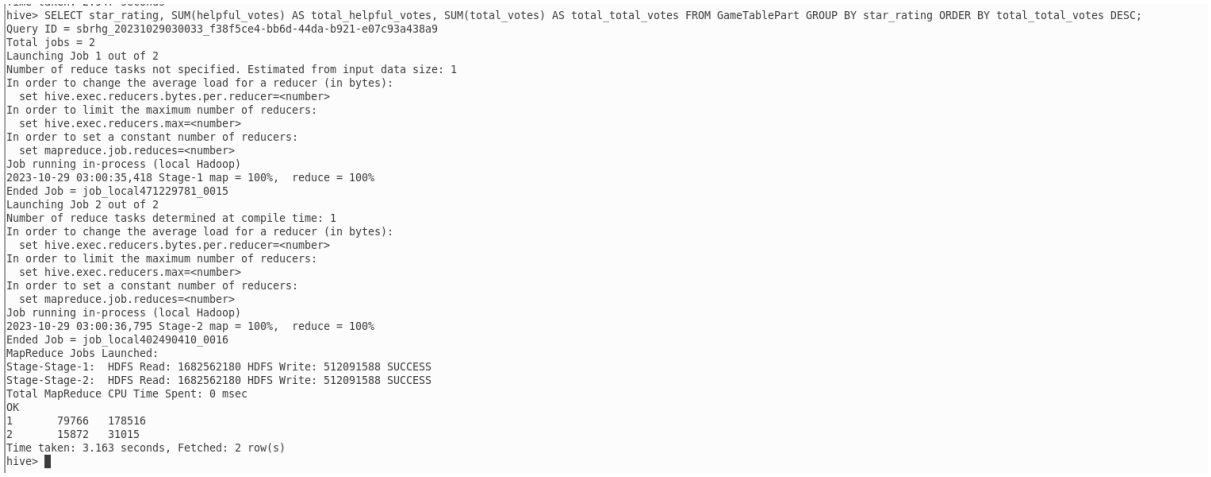
Insert Into GameTablePart Partition (star_rating = 1) Select marketplace,customer_id,review_id,product_id,product_parent,product_title,product_category,helpful_votes,total_votes,vine,verified_purchase,review_headline,review_body, review_date From GameTable Where star_rating=1;



Insert Into GameTablePart Partition (star_rating = 2) Select marketplace,customer_id,review_id,product_id,product_parent,product_title,product_category,hel

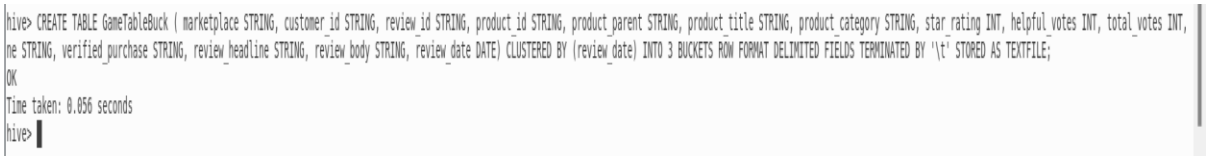
pful_votes,total_votes,vine,verified_purchase,review_headline,review_body, review_date From GameTable Where star_rating=2;

SELECT star_rating, SUM(helpful_votes) AS total_helpful_votes, SUM(total_votes) AS total_total_votes FROM GameTablePart GROUP BY star_rating ORDER BY total_total_votes DESC;



STUDENT-3

CREATE TABLE GameTableBuck (marketplace STRING, customer_id STRING, review_id STRING, product_id STRING, product_parent STRING, product_title STRING, product_category STRING, star_rating INT, helpful_votes INT, total_votes INT, vine STRING, verified_purchase STRING, review_headline STRING, review_body STRING, review_date DATE) CLUSTERED BY (review_date) INTO 3 BUCKETS ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS TEXTFILE;



INSERT INTO GameTableBuck SELECT * FROM GameTable;

```
hive> INSERT INTO GameTableBuck SELECT * FROM GameTable;
Query ID = sbrhg_20231029030324_cdfe908f-cb69-4168-b834-d1c0a7477318
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks determined at compile time: 3
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>
Job running in-process (local Hadoop)
2023-10-29 03:03:26,741 Stage-1 map = 0%,  reduce = 0%
2023-10-29 03:03:28,749 Stage-1 map = 100%,  reduce = 0%
2023-10-29 03:03:29,765 Stage-1 map = 100%,  reduce = 67%
2023-10-29 03:03:30,767 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local473275780_0017
Loading data to table default.gametablebuck
Launching Job 2 out of 2
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>
Job running in-process (local Hadoop)
2023-10-29 03:03:32,438 Stage-3 map = 100%,  reduce = 100%
Ended Job = job_local1485004021_0018
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 3657742200 HDFS Write: 1170445367 SUCCESS
Stage-Stage-3:  HDFS Read: 1828871100 HDFS Write: 658400640 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
Time taken: 7.663 seconds
hive>
```

SELECT MIN(review_date), MAX(review_date) FROM GameTableBuck WHERE INPUT__FILE__NAME LIKE '%000000_0';

```
hive> SELECT MIN(review_date), MAX(review_date) FROM GameTableBuck  WHERE INPUT__FILE__NAME LIKE '%000000_0';
Query ID = sbrhg_20231029030441_28aea450-94d9-428d-81ea-b50cb07aef6b
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>
Job running in-process (local Hadoop)
2023-10-29 03:04:43,187 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local429758708_0019
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 1975180152 HDFS Write: 658400640 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
2008-12-24      2015-08-30
Time taken: 1.867 seconds, Fetched: 1 row(s)
hive>
```

SELECT MIN(review_date), MAX(review_date) FROM GameTableBuck WHERE INPUT__FILE__NAME LIKE '%000001_0';

```
hive> SELECT MIN(review_date), MAX(review_date) FROM GameTableBuck  WHERE INPUT__FILE__NAME LIKE '%000001_0';
Query ID = sbrhg_20231029030525_99f74291-ddda-49e3-90ae-17109df1a693
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>
Job running in-process (local Hadoop)
2023-10-29 03:05:27,145 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local247412421_0020
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 2121488688 HDFS Write: 658400640 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
2006-08-08      2015-08-31
Time taken: 1.593 seconds, Fetched: 1 row(s)
hive>
```

SELECT MIN(review_date), MAX(review_date) FROM GameTableBuck WHERE INPUT__FILE__NAME LIKE '%000002_0';

```
hive> SELECT MIN(review_date), MAX(review_date) FROM GameTableBuck  WHERE INPUT__FILE__NAME LIKE '%000002_0';
Query ID = sbrhg_20231029030613_dd8ed2ab-1f69-45de-bad9-cbff1e9dcd25
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>
Job running in-process (local Hadoop)
2023-10-29 03:06:15,144 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local1583026834_0021
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 2267797224 HDFS Write: 658400640 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
2009-02-04      2015-08-29
Time taken: 1.709 seconds, Fetched: 1 row(s)
hive>
```

SELECT product_id, AVG(helpful_votes) AS avg_helpful_votes, AVG(total_votes) AS avg_total_votes FROM GameTable GROUP BY product_id HAVING AVG(helpful_votes) > 1;

```
hive> SELECT product_id, AVG(helpful_votes) AS avg_helpful_votes, AVG(total_votes) AS avg_total_votes FROM GameTable GROUP BY product_id HAVING AVG(helpful_votes) > 1;
Query ID = sbrhg_20231029030657_dd632a9d-d183-416e-b11f-af953fef746f
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>
Job running in-process (local Hadoop)
2023-10-29 03:06:59,668 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local337392563_0022
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 2414106144 HDFS Write: 658400640 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
B000AQ7K4I      3.5     3.5
B001AU6TQ8      2.923076923076923       3.3846153846153846
B001AUEITS      2.6470588235294117      3.0
B001JP5H26      3.217391304347826       4.695652173913044
B001K7HUR8      1.3269230769230769      2.3269230769230773
B001K7HUUU      1.2142857142857142      1.3571428571428572
B001K7HUWI      2.0     2.0
B001K7HUWS      2.0     2.125
B001K7HV0O      3.0     3.0
B001K7HV1S      1.6666666666666667      1.6666666666666667
B001K7HV2C      1.4     1.6
B001K7HV2M      1.0909090909090908      1.3636363636363635
B001KBZZU8      2.0     2.0
B001KBZZXA      1.25    1.25
B001KBZZXU      2.0434782608695654      2.4347826086956523
B001KBZZYO      2.0     4.0
B001KC000W      1.6666666666666667      1.6666666666666667
B001KC001G      4.0     5.0
B001KC0020      2.5     2.875
B001KC003Y      1.5     2.0
B001KC0070      2.75    3.0
B001KC008Y      2.0     2.0
B001KC00BQ      3.888888888888889       4.666666666666667
B001KC00C0      1.4166666666666667      1.5
B001KC00CK      1.263157894736842       2.245614035087719
B001KC00HA      3.0     3.230769230769231
B001KC00HU      1.728813559322034       4.169491525423729
B001KC00IO      2.8     4.5
```