

# HBase

## Overview:

Download amazon video game review data from the link [https://s3.amazonaws.com/amazon-reviews-pds/tsv/amazon\\_reviews\\_us\\_Video\\_Games\\_v1\\_00.tsv.gz](https://s3.amazonaws.com/amazon-reviews-pds/tsv/amazon_reviews_us_Video_Games_v1_00.tsv.gz) (Links to an external site.) . After extracting you will get a “.tsv” file where the columns are ‘tab’ separated. The description of the columns can be found in the link <https://s3.amazonaws.com/amazon-reviews-pds/tsv/index.txt> (Links to an external site.) . For this assignment, you must run different queries on **HBase** using this dataset.

### Part 1

- \* Download and insert this dataset to a HBase table named "GameTable":

- \* Keep three column families:

- (i) Info (contains columns 'marketplace' and 'verified\_purchase')
- (ii) Rating (contains columns 'star\_rating', 'helpful\_votes', 'total\_votes', )
- (iii) Review (contains columns 'review\_headline', 'review\_body')

The columns under 'Rating' should be integer. The columns under 'Info' and 'Review' should be string. Concat 'customer\_id' and 'review\_id' by an underscore to use it as the Row Key. (for example, some random Row Key for a certain row can be 11\_22, where 11 is the customer\_id and 22 is the review\_id. After creating the table and inserting data, run 'describe' command to show that the table has been created perfectly and then run 'scan' & 'limit' command to show 5 rows.

### Part 2

- \* Alter 'Review' column family to support 3 versions. Now, take a random row to put additional 2 different 'review\_body'. Then show the all 3 'review\_body' for this row.

- \* Find the 'review\_bodys' that have the word 'awesome'.

- \* Find the 'review\_headlines' that have any characters apart from alphanumeric characters (use regex).

### Part 3:

- \* Find how many reviews have 'star\_rating' equal to 5.

- \* Find the average 'helpful\_vote' in the dataset.

- \* Show the 'review\_headlines' that got 1 'star\_rating'.

## Part 1:

Download and Insert this dataset to a HBase table named "GameTable":



InsertData.java

Data Inserted: 1785997 rows

```
Applications Places System cqs-cs6304-119.ats.mst.edu Sat Oct
Java - HBase_Demo/src/InsertData.java - Eclipse
Edit Source Refactor Navigate Search Project Run Window Help
Problems Javadoc Declaration Console Debug
<terminated> InsertData [Java Application] /usr/java/jdk1.7.0_67-cloudera/bin/java (Oct 16, 2021, 2:08:48 PM)
21/10/16 14:08:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using
21/10/16 14:08:50 INFO zookeeper.RecoverableZooKeeper: Process identifier=hconnection-0x55247653 connecting
21/10/16 14:08:50 INFO zookeeper.ZooKeeper: Client environment:zookeeper.version=3.4.5-cdh5.12.0--1, built c
21/10/16 14:08:50 INFO zookeeper.ZooKeeper: Client environment:host.name=quickstart.cloudera
21/10/16 14:08:50 INFO zookeeper.ZooKeeper: Client environment:java.version=1.7.0_67
21/10/16 14:08:50 INFO zookeeper.ZooKeeper: Client environment:java.vendor=Oracle Corporation
21/10/16 14:08:50 INFO zookeeper.ZooKeeper: Client environment:java.home=/usr/java/jdk1.7.0_67-cloudera/jre
21/10/16 14:08:50 INFO zookeeper.ZooKeeper: Client environment:java.class.path=/home/cloudera/git/HBase_Demo
21/10/16 14:08:50 INFO zookeeper.ZooKeeper: Client environment:java.library.path=/usr/java/packages/lib/amd6
21/10/16 14:08:50 INFO zookeeper.ZooKeeper: Client environment:java.io.tmpdir=/tmp
21/10/16 14:08:50 INFO zookeeper.ZooKeeper: Client environment:java.compiler=<NA>
21/10/16 14:08:50 INFO zookeeper.ZooKeeper: Client environment:os.name=Linux
21/10/16 14:08:50 INFO zookeeper.ZooKeeper: Client environment:os.arch=amd64
21/10/16 14:08:50 INFO zookeeper.ZooKeeper: Client environment:os.version=2.6.32-754.el6.x86_64
21/10/16 14:08:50 INFO zookeeper.ZooKeeper: Client environment:user.name=cloudera
21/10/16 14:08:50 INFO zookeeper.ZooKeeper: Client environment:user.home=/home/cloudera
21/10/16 14:08:50 INFO zookeeper.ZooKeeper: Client environment:user.dir=/home/cloudera/git/HBase_Demo
21/10/16 14:08:50 INFO zookeeper.ZooKeeper: Initiating client connection, connectString=localhost:2181 sessi
21/10/16 14:08:50 INFO zookeeper.ClientCnxn: Opening socket connection to server localhost/127.0.0.1:2181. W
21/10/16 14:08:50 INFO zookeeper.ClientCnxn: Socket connection established, initiating session, client: /127
21/10/16 14:08:50 INFO zookeeper.ClientCnxn: Session establishment complete on server localhost/127.0.0.1:21
Inserted 1785997 Inserted
```

hbase(main):001:0> describe 'GameTable'

```
21/10/16 14:10:17,425 INFO [main] Configuration.deprecation: hadoop.native.li
b is deprecated. Instead, use io.native.lib.available
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.0-cdh5.12.0, rUnknown, Thu Jun 29 04:42:07 PDT 2017

hbase(main):001:0> describe 'GameTable'
Table GameTable is ENABLED
GameTable
COLUMN FAMILIES DESCRIPTION
{NAME => 'Info', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATIO
N_SCOPE => '0', VERSIONS => '1', COMPRESSION => 'NONE', MIN_VERSIONS => '0', TTL
=> 'FOREVER', KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_MEMORY
=> 'false', BLOCKCACHE => 'true'}
{NAME => 'Rating', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICAT
ION_SCOPE => '0', VERSIONS => '1', COMPRESSION => 'NONE', MIN_VERSIONS => '0', T
TL => 'FOREVER', KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_MEMORY
=> 'false', BLOCKCACHE => 'true'}
{NAME => 'Review', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICAT
ION_SCOPE => '0', VERSIONS => '1', COMPRESSION => 'NONE', MIN_VERSIONS => '0', T
TL => 'FOREVER', KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_MEMORY
=> 'false', BLOCKCACHE => 'true'}
3 row(s) in 0.3600 seconds
```



```
hbase(main):004:0> scan 'GameTable', {'LIMIT'=>5}
```

```
10000099_R38GX1BSE6I column=Rating:star_rating, timestamp=1634418631509, value=
3LM \x00\x00\x00\x05
10000099_R38GX1BSE6I column=Rating:total_votes, timestamp=1634418631509, value=
3LM \x00\x00\x00\x00
10000099_R38GX1BSE6I column=Review:review_body, timestamp=1634418631509, value=
3LM I love this charger. Works perfect and charges way faster
than the use method
10000099_R38GX1BSE6I column=Review:review_headline, timestamp=1634418631509, va
3LM lue=Five Stars
10000124_R312DHUSFZP column=Info:marketplace, timestamp=1634418967941, value=US
Q0P
10000124_R312DHUSFZP column=Info:verified_purchase, timestamp=1634418967941, va
Q0P lue=Y
10000124_R312DHUSFZP column=Rating:helpful_votes, timestamp=1634418967941, valu
Q0P e=\x00\x00\x00\x00
10000124_R312DHUSFZP column=Rating:star_rating, timestamp=1634418967941, value=
Q0P \x00\x00\x00\x02
10000124_R312DHUSFZP column=Rating:total_votes, timestamp=1634418967941, value=
Q0P \x00\x00\x00\x00
10000124_R312DHUSFZP column=Review:review_body, timestamp=1634418967941, value=
Q0P Sadly i was one of the people who got a bad unit. The char
ging port wasnt even charging my controller without me wat
ching it and holding it down just so the contact needles h
it the battery :\x5C
10000124_R312DHUSFZP column=Review:review_headline, timestamp=1634418967941, va
Q0P lue=Not worth the buy
5 row(s) in 0.2470 seconds
```

## Part 2:

```
hbase(main):017:0> alter 'GameTable',{'NAME'=>'Review','VERSIONS'=>3}
```

```
File Edit View Search Terminal Help
cqs-cs6304-119.ats.mst.edu
cloudera@quickstart:~

[cloudera@quickstart ~]$ hbase shell
2021-10-16 14:40:20,161 INFO [main] Configuration.deprecation: hadoop.native.li
b is deprecated. Instead, use io.native.lib.available
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.0-cdh5.12.0, rUnknown, Thu Jun 29 04:42:07 PDT 2017

hbase(main):001:0> describe 'GameTable'
Table GameTable is ENABLED
GameTable
COLUMN FAMILIES DESCRIPTION
{NAME => 'Info', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATIO
N_SCOPE => '0', VERSIONS => '1', COMPRESSION => 'NONE', MIN_VERSIONS => '0', TTL
=> 'FOREVER', KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_MEMORY =>
'false', BLOCKCACHE => 'true'}
{NAME => 'Rating', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICAT
ION_SCOPE => '0', VERSIONS => '1', COMPRESSION => 'NONE', MIN_VERSIONS => '0', T
TL => 'FOREVER', KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_MEMORY
=> 'false', BLOCKCACHE => 'true'}
{NAME => 'Review', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICAT
ION_SCOPE => '0', VERSIONS => '1', COMPRESSION => 'NONE', MIN_VERSIONS => '0', T
TL => 'FOREVER', KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_MEMORY
=> 'false', BLOCKCACHE => 'true'}
3 row(s) in 0.2750 seconds

hbase(main):002:0> alter 'GameTable',{'NAME'=>'Review','VERSIONS'=>3}
Updating all regions with the new schema...
0/2 regions updated.
1/2 regions updated.
2/2 regions updated.
Done.
0 row(s) in 4.5210 seconds
```

```
hbase(main):001:0> describe 'GameTable'
```

```
Applications Places System cos-cv6204-119.ats.mst.edu Sat Oct 16, 2:44 PM cloudera
cloudera@quickstart:~$
base(main):001:0> describe 'GameTable'
table GameTable is ENABLED
GameTable
COLUMN FAMILIES DESCRIPTION
NAME => 'Info', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATION_SCOPE => '0', VERSIONS => '1', COMPRESSION => 'NONE', MIN_VERSIONS => '0', TTL => 'FOREVER', KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_MEMORY => 'false', BLOCKCACHE => 'true'
NAME => 'Rating', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATION_SCOPE => '0', VERSIONS => '1', COMPRESSION => 'NONE', MIN_VERSIONS => '0', TTL => 'FOREVER', KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_MEMORY => 'false', BLOCKCACHE => 'true'
NAME => 'Review', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATION_SCOPE => '0', VERSIONS => '1', COMPRESSION => 'NONE', MIN_VERSIONS => '0', TTL => 'FOREVER', KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_MEMORY => 'false', BLOCKCACHE => 'true'
row(s) in 0.2750 seconds

base(main):002:0> alter 'GameTable', {'NAME'=>'Review','VERSIONS'=>3}
updating all regions with the new schema...
/2 regions updated.
/2 regions updated.
/2 regions updated.
one.
row(s) in 4.5210 seconds

base(main):003:0> describe 'GameTable'
table GameTable is ENABLED
GameTable
COLUMN FAMILIES DESCRIPTION
NAME => 'Info', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATION_SCOPE => '0', VERSIONS => '1', COMPRESSION => 'NONE', MIN_VERSIONS => '0', TTL => 'FOREVER', KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_MEMORY => 'false', BLOCKCACHE => 'true'
NAME => 'Rating', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATION_SCOPE => '0', VERSIONS => '1', COMPRESSION => 'NONE', MIN_VERSIONS => '0', TTL => 'FOREVER', KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_MEMORY => 'false', BLOCKCACHE => 'true'
NAME => 'Review', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATION_SCOPE => '0', VERSIONS => '3', COMPRESSION => 'NONE', MIN_VERSIONS => '0', TTL => 'FOREVER', KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_MEMORY => 'false', BLOCKCACHE => 'true'
row(s) in 0.0210 seconds

base(main):004:0>
```

## Random row to put additional 2 different 'review\_body':

hbase(main):017:0> put 'GameTable','10000124\_R312DHUSFZPQ0P','Review:review\_body', 'This value is for version2'

hbase(main):017:0> put 'GameTable','10000124\_R312DHUSFZPQ0P','Review:review\_body', 'This value is for version3'

hbase(main):017:0>

get 'GameTable','10000124\_R312DHUSFZPQ0P', {'COLUMNS'=>'Review:review\_body','VERSIONS'=>3}

```
hbase(main):009:0> get 'GameTable','10000124_R312DHUSFZPQ0P', {'COLUMNS'=>'Review:review_body','VERSIONS'=>3}
COLUMN CELL
Review:review_body timestamp=1634421154271, value=This value is for version3
Review:review_body timestamp=1634421133528, value=This value is for version2
Review:review_body timestamp=1634418967941, value=Sadly i was one of the people who got a bad unit. The charging port wasnt even charging my controller without me watching it and holding it down just so the contact needles hit the battery :(\x5C
3 row(s) in 0.0140 seconds

hbase(main):010:0>
```

## Substring: "awesome" 102690 rows

hbase(main):017:0>



Applications Places System Sat Oct 16, 3:04 PM cloudera

File Edit View Search Terminal Help

9998961\_RYSKQVTSPP51

column=Review:review body, timestamp=1634419471692, value=This is the best Call of duty game so far, since Modern Warfare 1. The game is better than MW2, graphics wise its the same, a little bit better tho. Multiplayer online its awesome, and this is coming from a gamer that played Call of Duty 4 on promod leagues competitively for 3 years. My PC consist on: Intel i5 750 CPU, Sapphire HD 5850, 4 GB G.Skill ram, Samsung 2233r LCD monitor, CoolerMaster HAF 922 case. Game running at 1680 x 1050, 4x aa, extra settings, getting 70-80 FPS on single player campaign, and 60 to 125 FPS(capped by console command) during online gameplay. Forgive all the lag and freezes madness, that was because they configured wrong the settings on the servers the first day, problem solved since day 2. In liking the game more and more every time i play, just waiting to the mod tools to be released, for some promod action and the game will be perfect!! PD: Do not take into account the noob hate r reviews of the game, if u like FPS and Call of Duty, buy the game, you wont regret it.

9999700\_R3DFBAG3LAYAQ

column=Review:review body, timestamp=1634419157414, value=Got it properly packaged no problems put it to use right away work s great does exactly what i advertised all and all a great item for price and quality is awesome.

9999834\_R3JBZ64NZQ85C

column=Review:review body, timestamp=1634418824104, value=I will begin this review saying that i normally don't write review s i read alot of them and this is my first writing one. I have played every forza game up to this point, starting with Forza Motorsport on the original xbox. The games were definitely fun in but i always wondered what would forza would be like if there was an open world. My dream became a reality when Forza horizon was released on the 360. I loved that game, it was definitely new and fresh for the forza series. However for me the game was short lived because their wasn't many things to do. As time went on and the xbox one was released i wondered again if a Forza horizon would be released and here we t oday.<br />-<br />Initial Impression: Upon starting the game, they give you this preview that for me sent chills down my spine e. You can literally close your eyes and the narrator will send you off into a deep thought about cars, racing and partying. After that is done they throw you in a huracan and off you go to the horizon festival. Let me tell you, the graphics on this s game are very nice. You can see reflections off the windshield, dirt as you drive off road and scratches off the things you u hit. Interiors are called are wonderful to look at as well. In addition to that, the sounds of the cars are really close t o how they are in reality.<br />-<br />Differences between it's younger brother: Forza horizon had a totally different scenery, considering the map was based off of Colorado. It was on the smaller side and there wasn't much free roam off the actual road. Here comes Horizon 2 with a map that is probably double the size of the original and based off of Italy and France. This map is a lot larger with plenty of different landscape that will blow your mind away. Also the car selection, in my opinion, is a lot better this time around. They have something for everyone. Plus you can fine tune your car, which was impossible e in the last games.<br />-<br />Gameplay: Gameplay is great. Lots of different races including sprint, circuit, cross country e, trail races and showcases. Modifying the cars in game is awesome. Painting them are fun as well. The road trips add a different sense into the game in where it doesn't feel like a video game anymore.<br />-<br />The bad: There are some sound glitch e in the game relating to the cars. One car doesn't have any sound at all and i have came across like one or two graphical e errors. Other then that i don't think anything else is bad.<br />-<br />Overall: The game is amazing. From the introduction, t o the cars, to the scenery, the developers did extremely well on this. It actually made me re-buy the Xbox one. If you can g et past the slight glitches in the game play, this game will keep you entertained for hours. If you want an even better time grab your friends and go for a road trip. Give the game a shot. If your a car guy like you, there is no reason why you shou ldn't like it.

102690 row(s) in 71.5370 seconds

hbase(main):011:0>

HBase\_Demo/Comma... Java - HBase\_Demo/sr... amazon\_reviews\_us... cloudera@quickstart:~ cloudera@quickstart:~

```
hbase(main):017:0>
```

```
Applications Places System
File Edit View Search Terminal Help
18224249 R13FO56FASRCML
286489 R1C13092AB999X
121498 R12P327M2G320C1
2391818 R1U20R08P5Q20Q
124604 R1C13092AB999X
3065925 R1D2P30Q0E0E0
454697 R1N02B4A4C7F52
454697 R1N02B4A4C7F52
454697 R1U10C17H138K3
454697 R1C13092AB999X
4555573 R1L18N7H93K04
4555573 R1C13092AB999X
552328 R1B6P47W6J87G1
552328 R1C13092AB999X
7058431 R1FCE44A2D2010
7058431 R1C13092AB999X
8661346 R1BFE165X1M93M
8661346 R1C13092AB999X
8775805 R1X51H8K5H93
8775805 R1C13092AB999X
8775805 R2017FAG77P4Y
8775805 R1C13092AB999X
9834490 R1Q20B0A997Y0K1
9834490 R1C13092AB999X
9987456 R2E42710ABL11N
9987456 R1C13092AB999X
9987456 R1WY02736A2JN
9987456 R1Y12248B178C
991335 R1EY5J36G2J8K
9914744 R1K103195502E0
9914744 R1E96A6G5701Y
9925998 R3OP6B71136U2
9925998 R1V5V6K0N0V2G
9972823 R205J62739H9U
9913855 R15J4YU102DHF
798 708 in 71.7600 seconds

column-Review: review headline, timestamp=1634419132779, value=...
121498 R12P327M2G320C1, timestamp=1634419636966, value=121498 R12P327M2G320C1
2391818 R1U20R08P5Q20Q, timestamp=1634419678074, value=2391818 R1U20R08P5Q20Q
3065925 R1D2P30Q0E0E0, timestamp=1634419667981, value=3065925 R1D2P30Q0E0E0
454697 R1N02B4A4C7F52, timestamp=1634419674080, value=454697 R1N02B4A4C7F52
454697 R1N02B4A4C7F52, timestamp=1634419673804, value=454697 R1N02B4A4C7F52
454697 R1U10C17H138K3, timestamp=1634419932549, value=454697 R1U10C17H138K3
454697 R1C13092AB999X, timestamp=1634419888478, value=454697 R1C13092AB999X
552328 R1B6P47W6J87G1, timestamp=1634419916794, value=552328 R1B6P47W6J87G1
552328 R1C13092AB999X, timestamp=1634419167904, value=552328 R1C13092AB999X
7058431 R1FCE44A2D2010, timestamp=1634419179804, value=7058431 R1FCE44A2D2010
7058431 R1C13092AB999X, timestamp=1634419197804, value=7058431 R1C13092AB999X
8661346 R1BFE165X1M93M, timestamp=1634419173347, value=8661346 R1BFE165X1M93M
8661346 R1C13092AB999X, timestamp=1634419627396, value=8661346 R1C13092AB999X
8775805 R1X51H8K5H93, timestamp=1634419733347, value=8775805 R1X51H8K5H93
8775805 R2017FAG77P4Y, timestamp=163441967516, value=8775805 R2017FAG77P4Y
8775805 R1C13092AB999X, timestamp=163441968659, value=8775805 R1C13092AB999X
9834490 R1Q20B0A997Y0K1, timestamp=163441967516, value=9834490 R1Q20B0A997Y0K1
9834490 R1C13092AB999X, timestamp=163441968659, value=9834490 R1C13092AB999X
9987456 R2E42710ABL11N, timestamp=163441967516, value=9987456 R2E42710ABL11N
9987456 R1C13092AB999X, timestamp=163441968659, value=9987456 R1C13092AB999X
9987456 R1WY02736A2JN, timestamp=163441968659, value=9987456 R1WY02736A2JN
9987456 R1Y12248B178C, timestamp=163441968659, value=9987456 R1Y12248B178C
991335 R1EY5J36G2J8K, timestamp=163441968659, value=991335 R1EY5J36G2J8K
9914744 R1K103195502E0, timestamp=163441968659, value=9914744 R1K103195502E0
9914744 R1E96A6G5701Y, timestamp=163441968659, value=9914744 R1E96A6G5701Y
9925998 R3OP6B71136U2, timestamp=163441968659, value=9925998 R3OP6B71136U2
9925998 R1V5V6K0N0V2G, timestamp=163441968659, value=9925998 R1V5V6K0N0V2G
9972823 R205J62739H9U, timestamp=163441968659, value=9972823 R205J62739H9U
9913855 R15J4YU102DHF, timestamp=163441968659, value=9913855 R15J4YU102DHF
798 708 in 71.7600 seconds

base(main):032:0
HBase Demo/0.1
[java HBase Demo/0.1] amazon-reviews-us cloudera@quickstart:~ cloudera@quickstart:~
```

### Part 3:

- ```
hbase(main):017:0> scan 'GameTable', {'COLUMNS'=>'Rating:star_rating',  
'FILTER'=>"SingleColumnValueFilter('Rating', 'star_rating', '=', 'binary:\x00\x00\x00\x05')"} }
```

```
Applications Places System
qgs-6304-119.ats.mst.edu
cloudera@quickstart:~$

File Edit View Search Terminal Help

9998706 R2FH0P57F04P2Q column=Rating:star rating, timestamp=1634418991012, value=\x00\x00\x00\x05
9998726 R3GV7MIH1JR2Y8 column=Rating:star rating, timestamp=1634419599989, value=\x00\x00\x00\x05
9998761 RTL4R8VUSZSK20 column=Rating:star rating, timestamp=1634418767835, value=\x00\x00\x00\x05
9998943 RIAT0VTACSN1760 column=Rating:star rating, timestamp=1634419442350, value=\x00\x00\x00\x05
9998954 R8G9Y3HTUSR1Y column=Rating:star rating, timestamp=1634419564964, value=\x00\x00\x00\x05
9998954 R3JKM4F8J30AL column=Rating:star rating, timestamp=1634419495426, value=\x00\x00\x00\x05
9998961 R1C9H05519100 column=Rating:star rating, timestamp=1634419567773, value=\x00\x00\x00\x05
9998961 R3J8HA1PSS1TYAN column=Rating:star rating, timestamp=1634419340844, value=\x00\x00\x00\x05
9998961 RY5KDQVTS1WT5 column=Rating:star rating, timestamp=1634419417692, value=\x00\x00\x00\x05
9998994 R21A8163R50042 column=Rating:star rating, timestamp=1634419333532, value=\x00\x00\x00\x05
9998912 R30I8ASDM00B55 column=Rating:star rating, timestamp=1634418782285, value=\x00\x00\x00\x05
9998934 R2M7YRDZBNL9W9 column=Rating:star rating, timestamp=1634418563224, value=\x00\x00\x00\x05
9999110 R169FG53TCTM02 column=Rating:star rating, timestamp=1634418808085, value=\x00\x00\x00\x05
9999185 R6V03JZ81LIG1 column=Rating:star rating, timestamp=1634419087369, value=\x00\x00\x00\x05
9999311 R127Z8X6LMG6G5 column=Rating:star rating, timestamp=1634418964925, value=\x00\x00\x00\x05
9999360 R1VEKHPW10A719 column=Rating:star rating, timestamp=1634418553727, value=\x00\x00\x00\x05
9999368 R3A55D35X8KR05 column=Rating:star rating, timestamp=1634418662472, value=\x00\x00\x00\x05
9999415 R18I0I3P4033AH column=Rating:star rating, timestamp=1634418918066, value=\x00\x00\x00\x05
9999427 RCINTU08023K70 column=Rating:star rating, timestamp=1634419862210, value=\x00\x00\x00\x05
9999443 R2Z8R2NU4DZ7TV column=Rating:star rating, timestamp=1634419134725, value=\x00\x00\x00\x05
9999505 R27MFP3ZWHJ520 column=Rating:star rating, timestamp=1634418714738, value=\x00\x00\x00\x05
9999512 R1ZYUCLL44500C column=Rating:star rating, timestamp=1634419339722, value=\x00\x00\x00\x05
9999529 RYCU4TY1VNS1L column=Rating:star rating, timestamp=1634418961641, value=\x00\x00\x00\x05
9999577 R2VY164449ACLL column=Rating:star rating, timestamp=1634418791449, value=\x00\x00\x00\x05
9999577 R3F0W257Y0Y89D column=Rating:star rating, timestamp=1634418790812, value=\x00\x00\x00\x05
9999628 R1TH0PYFPU01YM column=Rating:star rating, timestamp=1634419019536, value=\x00\x00\x00\x05
9999645 R2U0GAHU3T1LAW8 column=Rating:star rating, timestamp=1634418718923, value=\x00\x00\x00\x05
9999666 R2E2Z70D3342AP column=Rating:star rating, timestamp=1634418632667, value=\x00\x00\x00\x05
9999790 R30F8AG3J4VAP0 column=Rating:star rating, timestamp=1634419157414, value=\x00\x00\x00\x05
9999834 R31B9Z2N2Z6N2Z column=Rating:star rating, timestamp=163418824184, value=\x00\x00\x00\x05
9999872 R26N137YK6C699 column=Rating:star rating, timestamp=163418898621, value=\x00\x00\x00\x05
9999897 R10CNDM1HSUBA95 column=Rating:star rating, timestamp=1634418863963, value=\x00\x00\x00\x05
9999933 R1F5MEPEFGQTZ5 column=Rating:star rating, timestamp=1634418549196, value=\x00\x00\x00\x05
9999933 R24LKG1190DZED column=Rating:star rating, timestamp=1634418550772, value=\x00\x00\x00\x05
9999981 R167Y5X48E50RV column=Rating:star rating, timestamp=1634419489923, value=\x00\x00\x00\x05

1026924 row(s) in 95.3080 seconds

hbase(main):017>:scan 'GameTable', {'COLUMNS'=>['Rating:star_rating'], 'FILTER'=>'SingleColumnValueFilter('Rating', 'star_rating',=, 'binary:\x00\x00\x00\x05')'}
```

ii) Find the average 'helpful\_vote' in the dataset.



Summation.java

The screenshot shows the Eclipse IDE with a Java project. The main editor displays the file `InsertData.java`, which contains a `main` method. The code uses `Configuration`, `HTable`, `Scan`, and `ResultScanner` from the `org.apache.hadoop.hbase` package. It configures a `Configuration` object, creates an `HTable`, scans the table, and prints the sum of the values in the `bytes` column.

```

15  public static void main(String[] args) throws IOException {
16      Configuration config = HBaseConfiguration.create();
17      @SuppressWarnings("resource")
18      HTable htable = new HTable(config);
19
20      Scan sc = new Scan();
21      sc.addColumn(Bytes.toBytes("bytes"));
22      double sum = 0;
23      double count = 0;
24
25      ResultScanner rsc = htable.getScanner(sc);
26
27      for (Result rr = rsc.next(); rr != null; rr = rsc.next()) {
28          int num = Bytes.toInt(rr.getValue(Bytes.toBytes("bytes")));
29          sum = sum + num;
30          count++;
31      }
32
33      System.out.println("Count: " + count);
34      System.out.println("Sum: " + sum);
35      System.out.println("Average: " + sum/count);
36
37  }

```

The console window on the right shows the output of the program, which is:

```

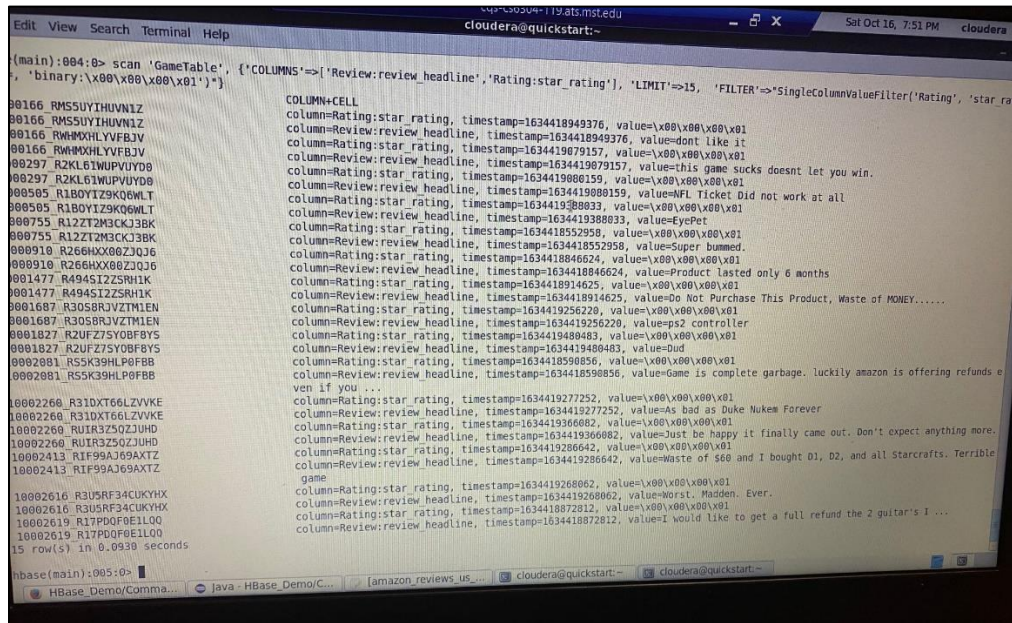
Count: 1785997.0
Sum: 4835191.0
Average: 2.2593492598251843

```



iii) Show the 'review\_headlines' that got 1 'star\_rating'

```
hbase(main):004:0> scan 'GameTable',  
{'COLUMNS'=>['Review:review_headline','Rating:star_rating'],'LIMIT'=>15,  
'FILTER'=>"SingleColumnValueFilter('Rating', 'star_rating', =, 'binary:\x00\x00\x00\x01')"}  
15 row(s) in 0.0939 seconds
```



The screenshot shows a terminal window with the following content:

```
hbase(main):004:0> scan 'GameTable', {'COLUMNS'=>['Review:review_headline','Rating:star_rating'],'LIMIT'=>15, 'FILTER'=>"SingleColumnValueFilter('Rating', 'star_rating', =, 'binary:\x00\x00\x00\x01')"}  
COLUMN=CELL  
00166 RMSSUYIHUVN1Z column=Rating:star_rating, timestamp=1634418949376, value=\x00\x00\x00\x01  
00166 RMSSUYIHUVN1Z column=Review:review_headline, timestamp=1634418949376, value=dont like it  
00166 RMSSUYIHUVN1Z column=Rating:star_rating, timestamp=1634419079157, value=\x00\x00\x00\x01  
00297 RZKL61WUPVYD0 column=Review:review_headline, timestamp=1634419079157, value=this game sucks doesnt let you win.  
00297 RZKL61WUPVYD0 column=Rating:star_rating, timestamp=1634419080159, value=\x00\x00\x00\x01  
00505 R180Y1ZK06MLT column=Review:review_headline, timestamp=1634419080159, value=NFL Ticket Did not work at all  
00505 R180Y1ZK06MLT column=Rating:star_rating, timestamp=1634419388033, value=\x00\x00\x00\x01  
00755 R12ZT2M3CJ3BK column=Review:review_headline, timestamp=1634418552958, value=EyePet  
00755 R12ZT2M3CJ3BK column=Rating:star_rating, timestamp=1634418552958, value=\x00\x00\x00\x01  
00910 R266HX00ZJQ36 column=Review:review_headline, timestamp=1634418846624, value=Super bunned.  
00910 R266HX00ZJQ36 column=Rating:star_rating, timestamp=1634418846624, value=\x00\x00\x00\x01  
00910 R266HX00ZJQ36 column=Review:review_headline, timestamp=1634418846624, value=Product lasted only 6 months  
00910 R266HX00ZJQ36 column=Rating:star_rating, timestamp=1634418914625, value=\x00\x00\x00\x01  
00910 R266HX00ZJQ36 column=Review:review_headline, timestamp=1634418914625, value=Do Not Purchase This Product, Waste of MONEY.....  
00910 R266HX00ZJQ36 column=Rating:star_rating, timestamp=1634419256220, value=\x00\x00\x00\x01  
00910 R266HX00ZJQ36 column=Review:review_headline, timestamp=1634419256220, value=ps2 controller  
00910 R266HX00ZJQ36 column=Rating:star_rating, timestamp=1634419480483, value=\x00\x00\x00\x01  
00910 R266HX00ZJQ36 column=Review:review_headline, timestamp=1634419480483, value=Dud  
00910 R266HX00ZJQ36 column=Rating:star_rating, timestamp=1634418598856, value=\x00\x00\x00\x01  
00910 R266HX00ZJQ36 column=Review:review_headline, timestamp=1634418598856, value=Game is complete garbage. luckily amazon is offering refunds e  
00910 R266HX00ZJQ36 column=Rating:star_rating, timestamp=1634419277252, value=\x00\x00\x00\x01  
00910 R266HX00ZJQ36 column=Review:review_headline, timestamp=1634419277252, value=AS bad as Duke Nukem Forever  
00910 R266HX00ZJQ36 column=Rating:star_rating, timestamp=1634419366882, value=\x00\x00\x00\x01  
00910 R266HX00ZJQ36 column=Review:review_headline, timestamp=1634419366882, value=Just be happy it finally came out. Don't expect anything more.  
00910 R266HX00ZJQ36 column=Rating:star_rating, timestamp=1634419286642, value=\x00\x00\x00\x01  
00910 R266HX00ZJQ36 column=Review:review_headline, timestamp=1634419286642, value=Waste of $60 and I bought 01, 02, and all Starcrafts. Terrible  
00910 R266HX00ZJQ36 column=Rating:star_rating, timestamp=1634419286642, value=\x00\x00\x00\x01  
00910 R266HX00ZJQ36 column=Review:review_headline, timestamp=1634419286642, value=Worst. Madden. Ever.  
00910 R266HX00ZJQ36 column=Rating:star_rating, timestamp=1634418872812, value=\x00\x00\x00\x01  
00910 R266HX00ZJQ36 column=Review:review_headline, timestamp=1634418872812, value=I would like to get a full refund the 2 guitar's I ...  
15 row(s) in 0.0939 seconds
```